

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
13 April 2006 (13.04.2006)

PCT

(10) International Publication Number  
**WO 2006/039405 A2**

(51) International Patent Classification:  
*C12Q 1/68* (2006.01) *G06F 19/00* (2006.01)

(21) International Application Number:  
PCT/US2005/035027

(22) International Filing Date:  
30 September 2005 (30.09.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/614,746 30 September 2004 (30.09.2004) US  
60/651,344 8 February 2005 (08.02.2005) US  
11/242,111 29 September 2005 (29.09.2005) US

(71) Applicant (for all designated States except US): **INTELLIGENESCAN, INC.** [US/US]; 3702 Autumn Glen Court, Santa Rosa, CA 95403 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **LEE, Nancy, M.** [US/US]; 1830 Funston Avenue, San Francisco, CA 94116 (US).

(74) Agents: **BASCH, Melissa, L.** et al.; Fliesler Meyer LLP, Four Embarcadero Center, Fourth Floor, San Francisco, CA 94111-4156 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

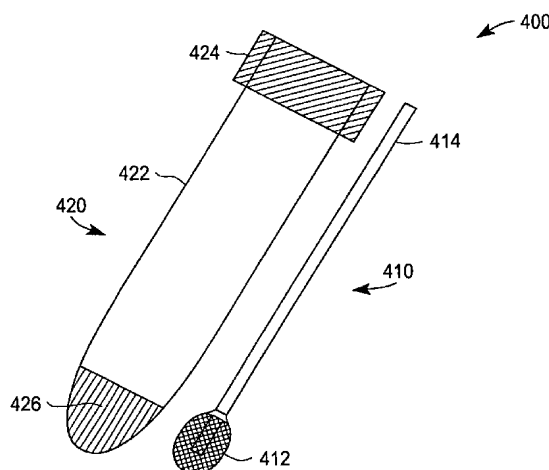
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: DRUG SCREENING AND MOLECULAR DIAGNOSTIC TEST FOR EARLY DETECTION OF COLORECTAL CANCER: REAGENTS, METHODS AND KITS THEREOF



(57) Abstract: A novel approach to the early detection of colorectal cancer ("CRC"), using a molecular diagnostic test to evaluate grossly normal-appearing colonic tissue for the early detection of colorectal cancer is disclosed. Such grossly normal-appearing colonic mucosal cells may be collected from non-invasive or minimally invasive procedures. The use of novel biomarker panels for drug screening also is disclosed. Such biomarker panels may be used wholly or in part as surrogate endpoints for monitoring effectiveness of a prospective drug in the intervention of pathologies, such as cancers, for example CRC, lung, prostate, and breast, and neurodegenerative diseases, for example Alzheimer's and ALS.



WO 2006/039405 A2



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**DRUG SCREENING AND MOLECULAR DIAGNOSTIC TEST FOR EARLY DETECTION OF COLORECTAL  
CANCER: REAGENTS, METHODS, AND KITS THEREOF**

5

**Claim of Priority**

U.S. Provisional Patent Application No. 60/614,746 entitled MOLECULAR  
DIAGNOSTIC TEST FOR EARLY DETECTION OF COLORECTAL CANCER: REAGENTS,  
METHODS, AND KITS THEREOF, by Nancy M. Lee, *et al.*, filed September 30, 2004  
(Attorney Docket No. NLEE-01001US0);

10

U.S. Provisional Patent Application No. 60/651,344 entitled METHODS OF USE OF  
A BIOMARKER PANEL FOR DRUG SCREENING, by Nancy M. Lee, *et al.*, filed February 8,  
2005 (Attorney Docket No. NLEE-01002US0); and

15

U.S. Patent Application No. 11/\_\_\_\_,\_\_\_\_ entitled DRUG SCREENING AND  
MOLECULAR DIAGNOSTIC TEST FOR EARLY DETECTION OF COLORECTAL  
CANCER: REAGENTS, METHODS, AND KITS THEREOF, by Nancy M. Lee, filed  
September 29, 2005 (Attorney Docket No. NLEE-01001US1).

**Cross-Reference to Related Applications**

20

This application is related to PCT/US2004/022594, entitled "Biomarker Panel for  
Colorectal Cancer," by Nancy M. Lee et al., filed July 14, 2004 (Attorney Docket No. NLEE-  
01000WO0), which claims priority to U.S. Provisional Application No. 60/488,660, entitled  
"Molecular Biomarker Panel for Determination of Colorectal Cancer," by Nancy M. Lee et al.,  
filed July 18, 2003 (Attorney Docket No. CPMC-01000US0), and also to U.S. Patent  
Application No. 10/690,880, entitled "Biomarker Panel for Colorectal Cancer," by Nancy M.  
Lee et al., filed October 22, 2003 (Attorney Docket No. CPMC-01000US1), each of which is  
incorporated herein in full, by reference.

25

Nucleotide and/or amino acid sequence listings are included in this application in  
computer-readable form and in hard-copy. The information included in computer-readable  
form is incorporated herein in full by reference. The information in computer-readable form  
is also included on diskette, and such information submitted on diskette is incorporated  
herein in full by reference. Compact diskette No. 1 contains the following file:  
NLEE1001WO0.ST25.txt (created 9/30/2005, 96K). The total number of diskettes submitted  
is one.

30

-2-

**Background**

The field of art of this disclosure concerns reagents, methods, and kits for the early detection of colorectal cancer ("CRC"), and methods for drug screening effective in the treatment of pathologies, such as cancers, for example, CRC, lung, prostate, and breast, and neurodegenerative diseases, for example Alzheimer's and ALS. These reagents, methods, and kits are based on a panel of biomarkers that are useful for risk assessment, early detection, establishing prognosis, evaluation of intervention, recurrence of CRC and other such pathologies, and drug discovery for therapeutic intervention.

In the field of medicine, clinical procedures providing for the risk assessment and early detection of CRC have been long sought. Currently, CRC is the second leading cause of cancer-related deaths in the Western world. One picture that has clearly emerged through decades of research into CRC is that early detection is critical to enhanced survival rates.

Thus, one long-sought approach for the early detection of CRC has been the search for biomarkers that are effective in the early detection of CRC, and therefore that are effective for the treatment of CRC. For more than four decades, since the discovery of carcinogenic embryonic antigen ("CEA"), the search for biomarkers effective for early detection of CRC has continued. It is further advantageous for sampling methods used in conjunction with an early diagnostic test for CRC to be minimally invasive or non-invasive. Non-invasive and minimally invasive sampling methods increase patient compliance, and generally reduce cost. Additionally, bioinformatic methods for analysis of complex, multivariate data typical of bioanalysis, yielding a reliable diagnostic evaluation based on such data sets, are also desirable.

Therapeutic intervention for numerous types of cancers, such as CRC, lung, prostate, and breast, includes surgery, chemotherapy, and radiation treatment, and combinations thereof. For CRC, a current area of continued research and development, in addition to search for non-invasive methods for early detection, is in the area of drug development.

One picture that has clearly emerged through decades of research into CRC is that early detection, coupled with effective therapeutic intervention is critical to enhanced survival rates. To date, the most commonly used drug in the treatment of CRC is 5-fluoruracil ("5FU"), which frequently is administered intravenously, in combination with the folic acid vitamin, leucovorin. A strategy referred to as primary chemotherapy is used when metastasis has occurred, and the cancer has spread to different parts of the body. For CRC, the current strategy for primary chemotherapy is the administration of an oral form of

-3-

5FU, capecitabine, in combination with Camptosar, a topoisomerase I inhibitor, or Eloxatin, an organometallic, platinum-containing drug that inhibits DNA synthesis.

Currently, strategies for new drug development for CRC include two areas of research: angiogenesis inhibitors, and signal transduction inhibitors.

5        Novel biopharmaceutical drugs include both protein- and ribozyme-based therapeutics. Humanized antibody-based therapeutics include examples such as Erbitux and Avastin. Erbitux, a signal transduction inhibitor, is aimed at inhibiting epidermal growth factor receptors ("EGFR") on the surface of cancerous cells. Avastin, an angiogenesis inhibitor, is aimed at inhibiting vascular endothelial growth factor ("VEGF"), which is known  
10      to promote the growth of blood vessels. Additionally, Angiozyme, an example of a ribozyme-based therapeutic, is an angiogenesis inhibitor directed against the expression of the VEGF-R1 receptor. New traditional small molecule-based drugs include examples such as Iressa, based on a quinazoline template, and acting as a signal transduction inhibitor, and SU11248, based on an indolinone template, which acts as an anti-angiogenesis  
15      inhibitor.

Still, a number of potential drawbacks and uncertainties remain for these nascent drug therapies for CRC. In addition to typical contraindications such as nausea, vomiting, headache, and diarrhea, other more serious side effects, such as gastrointestinal perforation, elevated or lowered blood pressure, extreme fatigue, and internal bleeding have  
20      been observed for many of the promising candidates. Additionally, though many of the drug therapies based on angiogenesis inhibition or signal transduction inhibition appear promising, they are in the very early stages of clinical trials.

Accordingly, a need exists in the art for biomarkers that are effective in the early detection of CRC, coupled with sampling methods that are minimally or non-invasive, and  
25      bioinformatic methods, which together produce a robust diagnostic test for the early detection of CRC. A need also exists in the art for drug development, which can provide effective treatment prior to the development of cancer for individuals diagnosed with pathologies, such as cancers, for example CRC, lung, prostate, and breast, and neurodegenerative diseases, for example Alzheimer's and ALS, while minimizing serious  
30      side effects.

#### **Brief Description of Figures**

**Fig. 1** is a table listing an embodiment of sequence listings for a panel of biomarkers of the disclosed invention.

-4-

**Fig. 2** is a distribution plot of control subjects versus test subjects evaluated using an aspect of the panel of biomarkers of **Fig. 1**, and an aspect of a bioinformatic evaluation of the disclosed invention.

**Fig. 3** shows the distribution of the log (base2) expression values for genes, PPAR- $\gamma$ , IL-8, SAA 1 and COX-2 and their cut-off points.

**Figs. 4A and 4B** show that expression of different genes is altered at different sites of MNCM from individuals with a family history of colon cancer.

**Fig. 5** displays a flow diagram of an aspect of the bioinformatic process used for evaluating data.

**Fig. 6** is an embodiment of a swab sampling and transport system for the minimally invasive sampling of colonic mucosal cells.

**Fig. 7** is a flow chart depicting one aspect of the drug screening disclosure.

**Fig. 8** is a flow chart depicting another aspect of the drug screening disclosure.

15

#### **Detailed Description**

To date, a greater understanding of the biology of CRC has been gained through the research on adenomatous polyposis coli ("APC"), p53, and Ki-ras genes, as well as the corresponding proteins, and related pathways involved regulation thereof. However, there is a distinct difference between research on a specific gene, its expression, protein product, and regulation, and understanding what genes are critical to include in a panel used for the analysis of CRC that is useful in the management of patient care for the disease. Panels that have been suggested for CRC are comprised of specific point mutations of the APC, p53, and Ki-ras, as well as BAT-26, which is a gene that is a microsatellite instability marker.

For CRC, biomarkers for risk assessment and early detection of CRC long have been sought. The difference between risk assessment and early detection is the degree of certainty regarding acquiring CRC. Biomarkers that are used for risk assessment confer less than 100% certainty of CRC within a time interval, whereas biomarkers used for early detection confer an almost 100% certainty of the onset of the disease within a specified time interval. Risk factors may be used as surrogate end points for individuals not diagnosed with cancer, providing that there is an established relationship between the surrogate end point and a definitive outcome. An example of an established surrogate end point for CRC is the example of adenomatous polyps. What has been established is that the occurrence of adenomatous polyps is a necessary, but not sufficient condition for an individual later to develop CRC. This is demonstrated by the fact that 90% percent of all preinvasive

-5-

cancerous lesions are adenomatous polyps or precursors, but not all individuals with adenomatous polyps go on later to develop CRC.

Adenomatous polyps have been established as surrogate end points for CRC, and adenomatous polyps are macroscopically identifiable by colonoscopy or sigmoidoscopy.

5 During such invasive procedures, biopsy samples can be taken from polyps or lesions for histological evaluation of the tissue. The molecular diagnostic approach disclosed herein may be used on grossly normal-appearing colonic mucosal cells that are not from a macroscopically identifiable polyp or lesion. However, as further disclosed herein, an invasive procedure need not be used to obtain a patient sample for histological evaluation.

10 A non-invasive or minimally-invasive procedure can be employed to obtain, for example, a blood sample, stool sample, or swab of grossly normal-appearing rectal cells, upon which a molecular diagnostic test can be performed to evaluate the presence or absence of CRC. No previously-described approach for early detection of CRC has disclosed the non-invasive or minimally invasive collection of grossly normal-appearing colonic mucosal cells (biopsy or

15 swab of rectal cells), blood samples, and/or stool samples, followed by a molecular and/or protein expression diagnostic test, which can detect changes in the tissue before any untoward histological changes indicating CRC are manifest.

**Fig. 1** is a table that gives an overview of the sequence listings included with this disclosure. The table of **Fig. 1** lists a panel of biomarkers useful in practicing the disclosed

20 invention. One embodiment of a biomarker panel is the 16 identified coding sequences given by SEQ. ID NOs 1-16, while another embodiment of a biomarker panel is the 16 identified proteins given by SEQ. ID NOs 17-32. These two embodiments represent molecular marker panels that provide the selectivity and sensitivity necessary for the early detection of CRC. It is to be understood that fragments and variants of the biomarkers

25 described in the sequence listings are also useful biomarkers in embodiments of panels used for the early detection of CRC. What is meant by fragment is any incomplete or isolated portion of a polynucleotide or polypeptide in the sequence listing. Further, it is recognized that almost daily, new discoveries are announced for gene variants, particularly for those genes under intense study, such as genes implicated in diseases like cancer.

30 Therefore, the sequence listings given are exemplary of what now is reported for a gene, but it is recognized that for the purpose of an analytical methodology, variants of the gene and their fragments also are included.

In **Fig. 1**, the entries 1-16 in the table are one aspect of a panel of biomarkers, which are polynucleotide coding sequences, and include the name and abbreviation of the gene.

35 Entries 17-32 in **Fig. 1** are another embodiment of a panel of biomarkers, which are protein,

-6-

or polypeptide, amino acid sequences that correspond to the coding sequences for entries 1-16. A biomarker, as defined by the National Institutes of Health ("NIH") is a molecular indicator of a specific biological property; a biochemical feature or facet that can be used to measure the progress of disease or the effects of treatment. A panel of biomarkers is a selection of biomarkers, which taken together can be used to measure the progress of disease or the effects of treatment. Biomarkers may be from a variety of classes of molecules. As previously mentioned, there remains a need for biomarkers for CRC having the selectivity and sensitivity required to be effective for early detection of CRC. Therefore, one embodiment of what is disclosed herein is the selection of an effective set of biomarkers that is differentiating in providing the basis for early detection of CRC.

In one aspect of this disclosure, for the early detection of CRC, expression levels of polynucleotides indicated as SEQ. ID NOs 1-16 are determined from cells in samples taken from patients by non-invasive or minimally invasive methods. The contemplated methods include blood sampling, stool sampling, and rectal cell swabbing or biopsy. Such analysis of polynucleotide expression levels frequently is referred to in the art as gene expression profiling. For gene expression profiling, levels of mRNA in a sample are measured as a leading indicator of a biological state -- in this case, as an indicator of CRC. One of the most common methods for analyzing gene expression profiling is to create multiple copies from mRNA in a biological sample (said sample taken from a patient as disclosed above, by non- or minimally-invasive methods) using a process known as reverse transcription. In the process of reverse transcription, the mRNA from the sample is isolated from cells in the biological sample, by methods well-known in the art. The mRNA then is used to create copies of the corresponding DNA sequence from which the mRNA was originally transcribed. In the reverse transcription amplification process, copies of DNA are created without the regulatory regions in the gene (*i.e.*, introns). These multiple copies made from mRNA are therefore referred to as "cDNA," which stands for complementary, or copy DNA. Entries 33-64 are the sets of primers that can be used in the reverse transcription process for each biomarker gene listed in entries 1-16. All nucleotide and amino acid biomarker sequences identified in SEQ. ID NOs 1-64 are found in a printout attached and included as subject matter of this application, and are found on a diskette also included as part of this application and incorporated herein by reference.

Since the reverse transcription procedure amplifies copies of cDNA proportional to the original level of mRNA in a sample, it has become a standard method that allows the identification and quantification of even low levels of mRNA present in a biological sample.

-7-

Genes either may be up-regulated or down-regulated in any particular biological state, and hence mRNA levels shift accordingly.

In one aspect of this disclosure, a method for gene expression profiling comprises the quantitative measurement of cDNA levels for at least two of the biomarkers of the panel of biomarkers selected from SEQ. ID NOs. 1-16, in a biological sample taken from a patient by a non- or minimally-invasive procedure, such as blood sampling, stool sampling, rectal cell swabbing, and/or rectal cell biopsy. The tissue taken need not be apparently diseased; in fact, the disclosed invention is contemplated to be useful in evaluating even grossly normal-appearing cells for detection of CRC. Such a method for gene expression profiling requires the use of primers, enzymes, and other reagents for the preparation, detection, and quantifying of cDNAs. The method of creating cDNA from mRNA in a sample is referred to as the reverse transcriptase polymerase chain reaction ("RT-PCR"). The primers listed in SEQ. ID NOs 33-64 are particularly suited for use in gene expression profiling using RT-PCR based on the disclosed biomarkers in the biomarker panel. A series of primers were designed using Primer Express Software (Applied Biosystems, Foster City, CA). Specific candidates were chosen, and then tested to verify that only cDNA was amplified, and not contaminated by genomic DNA. The primers listed in SEQ. ID NOs 33-64 were specifically designed, selected, and tested accordingly.

The primers listed in SEQ. ID NOs 33-64 are important in the step subsequent to creating cDNA from isolated cellular RNA, for quantitatively amplifying copies in the real time PCR of gene expression products of interest. Optimal primer sequence, and optimal primer length are key considerations in the design of primers. The optimal primer sequence may impact the specificity and sensitivity of the binding of the primer with the template. A primer length between 18-30 bases is considered an optimal range. Theoretically, 18 bases is the minimal length representing a unique sequence, which would hybridize at only one position in most eukaryotic genomes. The primers listed in SEQ. ID NOs 33-64 range in primer length between 21-27 bases, and were designed and validated to amplify cDNA for the panel of nucleotides selected from SEQ. ID NOs 1-16. The specificity of the primers was demonstrated by a single product on 10% polyacrylamide gel electrophoresis ("PAGE"), and a single dissociation curve of the PCR product.

Once the primer pairs have been designed, and validated for specificity, they may be synthesized in large quantities, and stored for convenient future use. Since the PCR reaction is sensitive to buffer concentration and buffer constituents, primers should be maintained in a suitable diluent that will not interfere in the amplification reaction. One example of a suitable diluent is 10 mM Tris buffer, with or without 1mM EDTA, depending on

-8-

the assay sensitivity to EDTA. Alternatively, another example of a suitable diluent for the primers is deionized water that is nuclease-free. The primers may be aliquoted in appropriate containers, such as siliconized tubes, and lyophilized if so desired. The liquid or lyophilized samples are preferably stored at refrigeration temperatures defined as long-term for biological samples, which is between about -20C° to about -70C°. The concentration of primer in the amplification reaction is typically between 0.1 to 0.5  $\mu$ M. The typical dilution factor from the stock solution to the final reaction mixture is about 10 times, so that the aliquoted stock solution of the primers is typically between about 1 and 5  $\mu$ M.

In addition to the specifically designed primers listed in SEQ. ID Nos. 33-64, reagents such as one including a dinucleotide triphosphate mixture having all four dinucleotide triphosphates (e.g., dATP, dGTP, dCTP, and dTTP), one having the reverse transcriptase enzyme, and one having a thermostable DNA polymerase, are required for RT-PCR. Additionally buffers, inhibitors, and activators also are required for the RT-PCR process.

**Fig. 2** depicts one aspect of a bioinformatic data reduction process used for the early detection of CRC, showing a distribution of Mahalanobis distance for 17 controls (left), compared with 14 individuals with family history of CRC (middle), and 24 individuals with polyps (right). Tissue samples taken from grossly normal-appearing colonic mucosal tissue were evaluated using the biomarker panel of polynucleotides selected from SEQ. ID NOs. 1-16. The means for the gene expression levels for each of the 16 genes represented by polynucleotides selected from SEQ. ID NOs 1-16 for each control and test subject were calculated in log base 2 domain. The multivariate means, in a 16 dimensional hyperspace, were then determined for the controls, based on a multivariate normal distribution, in order to establish limits of normal expression levels. For each control, the Mahalanobis distance ("M-dist") from the multivariate mean of the other 16 controls was measured, while the M-dist for each of the test subjects was determined from the multivariate mean of the 17 controls. In each group displayed in **Fig. 2**, all the biopsies from a single individual form a vertical row. For the individuals with polyps, astericks mark the biopsies from individuals with hyperplastic polyps. The horizontal line indicates the 95th percentile of a chi-square distribution with 16 degrees of freedom. All values above this line (corresponding to an M-dist of about 25) are different from the mean of controls at a level of  $p < 0.05$ . The data presented clearly show that there is an altered gene expression pattern in grossly normal colonic mucosal tissue samples for the test subjects. The data accordingly demonstrate the enhanced sensitivity and selectivity of a diagnostic test using the biomarker panel of polynucleotides selected from SEQ. ID NOs. 1-16.

**Fig. 3** displays a flow diagram **300** of an aspect of the bioinformatic process used for evaluating the data from samples analyzed using expression profiling of polynucleotides selected from SEQ. ID Nos. 1-16. The goal of the bioinformatic analysis used to analyze the gene expression data for the molecular diagnostic test using the panel of polynucleotides selected from SEQ. ID NOs 1-16 was to use a single, easy-to-calculate measure of abnormality. It is desirable to analyze expression patterns of all genes in the panel selected from SEQ. ID NOs 1-16 by multivariate analysis, since multivariate analysis determines the significance of changes of all expression levels, taken together. There are several kinds of multivariate tests which may be useful for the bioinformatic analysis used to assess the presence or absence of colorectal cancer in patient samples tested using the molecular diagnostic test disclosed herein. Examples of multivariate analysis tests useful in the assessment of data from patient samples tested using the panel of polynucleotide biomarkers selected from SEQ. ID NOs 1-16 include the ANOVA and the Mahalanobis distance ("M-Dist") tests.

ANOVA is a global test that accounts for correlations among expression levels. It is desirable for the multivariate ANOVA tests to be based on Wilks' lambda criterion and to be carried out on log(base 2) values for the data obtained using the molecular diagnostic test using the panel of polynucleotides selected from SEQ. ID NOs 1-16 to achieve normal distribution of values.

M-dist analysis is another example of a multivariate analysis that summarizes, in a single number, the differences between two patterns of gene expression, taking into account variability of each gene's expression and correlations among pairs of genes. M-dist is often used as a test for outliers (individual cases that are significantly different from all other individual cases in the group) in multivariate data. M-dist can be converted to p-values by reference to a chi-square distribution with degrees of freedom equal to the number of variables (*i.e.*, genes). However, to avoid reliance on an assumption of multivariate normality, it is desirable to compare M-dist for individual cases (*i.e.*, those with polyps) to controls using a rank sum test, the Mann-Whitney test. By using the Mann-Whitney analysis, the inferences concerning differences in expression patterns do not depend on the assumption of multivariate normality. Therefore, this method allows the determination of the significance of all the experimental subjects' expression levels taken together, as well as the significance of each individual expression value.

A working example of the foregoing disclosure is provided below. Hao, C-Y, et al., *Alteration of Gene Expression in Macroscopically Normal Colonic Mucosa from Individuals with a Family History of Sporadic Colon Cancer*, 11 Clin. Cancer Res., 1400-07 (Feb. 15,

-10-

2005). The example presented is provided as a further guide to the practitioner of ordinary skill in the art, and is not to be construed as limiting the invention in any way.

This example was undertaken to investigate whether expression of several genes was altered in morphologically normal colonic mucosa ("MNCM") of individuals who have not developed colon cancer, but are at high risk of doing so because of a family history of CRC.

#### Human subjects

Biopsies of MNCM from the rectum and sigmoid colon were performed at the time of routine colonoscopy from individuals seen at the California Pacific Medical Center ("CPMC") who had no history of prior colon cancer, and who were free of adenomatous polyps, colon cancer or other colonic lesions at the time of examination. Twelve individuals with a family history of colon cancer in a first-degree relative (Table 3) and sixteen individuals with no known family history of colon cancer were included in the study. Although the information of family cancer history is obtained by patients' self-reports without confirmation from the hospital's cancer registry, a recent study has confirmed the accuracy of self-reported family history with regard to colon cancer. Of the twelve individuals with a family history of colon cancer, two are mother and daughter (cases #6 and 7 in Table 3), two are sister and brother (cases #11 and 12), and the rest are not related. Study subjects ranged in age from 18 to 64 years in the group with a family history of colon cancer, and 16 to 83 years in the control group (the 16-year-old had undergone colonoscopy for chronic abdominal pain). The research protocols for obtaining normal biopsy specimens for study were approved by the CPMC Institutional Review Board. The appropriate procedure for obtaining informed consent was followed for all study subjects.

#### Extraction and preparation of RNA and cDNA

Biopsy samples obtained from the segment of colon between the cecum and the hepatic flexure were classified as ascending colon samples; those from the segment of colon between the hepatic flexure and the splenic flexure as transverse colon samples; those from the segment of colon below the splenic flexure as descending colon; those from the winding segment of colon below the descending colon were classified as rectosigmoid colon samples (approximately 5-25 cm from rectum). The number of biopsy samples obtained from each patient varied. Two to eight biopsy samples were obtained from each colon segment, except that only one sample was obtained from the transverse and the descending colon segments in one subject of the family history group. A total of 39 ascending colon, 37 transverse colon, 45 descending colon and 77 rectosigmoid specimens were obtained from the 12 individuals with a family history of colon cancer; and a total of 53 ascending colon, 48 transverse colon, 49 descending colon and 104 rectosigmoid

-11-

specimens were obtained from the 16 individuals with no family history of colon cancer. All biopsy samples were snap-frozen on dry ice and taken immediately to the laboratory for RNA preparation and reverse transcription as described.

#### Analysis of gene expression

5           The expression levels of oncogene c-myc, CD44 antigen ("CD44"), cyclooxygenase 1 and 2 ("COX-1" and "COX-2"), cyclin D1, cyclin-dependent kinase inhibitor ("p21<sup>cip/waf1</sup>"), interleukin 8 ("IL-8"), interleukin 8 receptor ("CXCR2"), osteopontin ("OPN"), melanoma growth stimulatory activity ("Gro $\alpha$ /MGSA"), GRO3 oncogene ("Gro $\gamma$ "), macrophage colony stimulating factor 1 ("MCSF-1"), peroxisome proliferative activated receptor, alpha, delta and  
10       gamma ("PPAR- $\alpha$ ,  $\delta$  and  $\gamma$ ") and serum amyloid A 1 ("SAA 1") were analyzed by quantitative RT-PCR. Quantitative RT-PCR were carried out. In brief, the cycle numbers ("C<sub>T</sub> value") were recorded when the accumulated PCR products crossed an arbitrary threshold. To normalize this value, a  $\Delta C_T$  value was determined as the difference between the C<sub>T</sub> value for each gene tested and the C<sub>T</sub> value for  $\beta$ -actin. The average  $\Delta C_T$  value for each gene in  
15       the control group was calculated. The  $\Delta\Delta C_T$  value was determined as the difference between the  $\Delta C_T$  value for each individual sample and the average  $\Delta C_T$  value for this gene obtained from the control samples. These  $\Delta\Delta C_T$  values were then used to calculate relative gene expression values as described. (Applied Biosystems, User Bulletin #2, December 11, 1997). All PCR were performed in duplicate when cDNA samples were available. The  
20       results were also verified using histidyl-tRNA synthetase as internal control. Relative gene expression values yielded similar results using either  $\beta$ -actin or his-tRNA synthetase as a reference. Statistical analyses reported here were obtained using  $\beta$ -actin as normalization controls.

#### Statistical analysis

25           Gene expression patterns were compared between individuals with a family history of colon cancer and the control group subjects who had no family history of colon cancer. Rather than testing expression of each gene separately and adjusting for multiple comparisons by methods that reduce statistical power, we tested the expression patterns of all genes by multivariate analysis of variance ("MANOVA") with Wilks' lambda criterion. This  
30       test is a multivariate analog of the F-test for univariate analysis of variance, which tests the equality of means. This type of analysis takes into account correlations among gene expression levels and controls the false-positive rate by providing a single test of whether the expression patterns, based on all the genes in the subset, differ between groups.

          If there was evidence that expression patterns differed between groups, we used  
35       univariate t-tests to determine which genes were contributing to the global difference. All

MANOVA tests were based on the Wilks' lambda criterion and were carried out on log (base 2) of the expression levels, since this transformation was required to achieve normal distributions. Our data consisted of a variable number of samples per subject with different numbers of individuals per group (family history vs. no family history). The analysis included

5 random effects terms for individuals within group and for samples within individuals to account for the sampling scheme. If  $Y_{ijk}$  denotes a log2 gene expression value for the  $k^{\text{th}}$  sample from the  $j^{\text{th}}$  patient from the  $i^{\text{th}}$  group, the statistical model is described mathematically by the equation:  $Y_{ijk} = M + A_i + B_{ij} + e_{ijk}$ , where  $A_i$  is the (fixed) group effect,  $B_{ij}$  is the (random) patient effect, and  $e_{ijk}$  is the (random) sample within patient effect.

10 We also tested whether or not the magnitude of the differential expression (over or under expression) increased along the colon from the ascending portion toward rectum, by defining a variable with value 1 for samples from the ascending, 2 for samples from the transverse, 3 for samples from the descending and 4 for samples from the rectosigmoid portion of the colon. This variable was added to the model so that its effect could be tested

15 for certain genes using univariate ANOVA.

#### Definition of cut-off point

The log (base 2) of the expression levels of all the biopsy samples from the control group was used to calculate the cut-off point for either up-regulation or down regulation of each gene. A table of tolerance bounds for a normal distribution was used to define cut-off

20 points so that a fraction of the distribution of no more than  $P$  would lie above the cut-off point for up-regulated genes or below the cut-off point for down-regulated genes. Each cut-off point was defined by cut-off point = mean +  $k(\text{SD})$ , where the mean and SD (Standard Deviation) are based on values from the control group. Values of  $k$  are found in the table and depend on the  $P$  value and the number of normal samples. Owen, D.B., Noncentral  $t$

25 and tolerance limits, in Brimbaum ZW, ed. Handbook of Statistical Tables, Reading, MA: Addison-Wesley, 1962, 108-127. Assuming a Gaussian distribution of expression levels of each gene, one would expect less than 1% of the biopsies from a normal population to have an expression level exceeding the 99% tolerance limit ( $p = 0.01$ ).

To calculate the probability that the number of observed samples outside the upper

30 99 percentile was due to chance in each case, we used the binomial distribution method with  $p = 0.01$  and  $n$  = the number of samples for each case multiplied by the number of genes tested. For example, for case #1 (Table 3) we had 2 samples; both showed abnormal expression for PPAR- $\gamma$  and SAA1, one of two for PPAR- $\delta$  and neither was abnormal for IL-8 and COX-2. Thus, for this case, 5 of 10 tested were beyond the upper

-13-

0.01 boundary. The probability that this happened by

chance is  $2.4 \times 10^{-8}$ . The general formula is given by:  $\Pr\{x \geq k \mid p, n\} = \sum_{i=k}^{5n} (0.01)^i (0.99)^{5n-i}$

5 where  $k$  is the number beyond the 99 percentile and  $n$  is the number of samples (5 is the number of genes tested).

### Results

Altered gene expression in the rectosigmoid mucosa of individuals with a family history of colon cancer:

10 Twelve individuals (ten women and two men) comprised the group with a family history of colon cancer; 16 individuals (nine women and seven men) served as the control group. (Table 1.) We analyzed a total of 92 ascending colon biopsy samples, 85 transverse colon samples, 94 descending colon biopsy samples and 181 rectosigmoid biopsy samples for levels of expression of 16 genes. Expressions of these genes are known to be altered in  
15 the late stages of human colon cancers. We have also shown that some of these genes are altered in the MNM from surgical resections of colon cancer patients.

Continuing to refer to Table 1, results represent analysis of 104 biopsy samples from the 16 individuals without family history and 77 biopsy samples from 12 individuals with family history of colon cancer in a first-degree relative. Samples were analyzed for  
20 gene expression as described in Methods. The numbers in the table represent the expression level relative to the average  $MC_T$  of the control group. If there is no variation among individuals, the normal gene expression level in the control group should equal to 1. Multivariate analysis using the Wilks Lambda criterion was carried out on log2 expression values of the 16 genes to determine the significance of the difference between  
25 the two groups. Genes are listed from smallest to largest  $P$  value.

Multivariate analysis of the expression values of all 16 genes indicated a significant difference in the biopsy samples from the rectosigmoid region ( $p = 0.01$ ) between those with and those without a family history of sporadic colon cancer. Gene expression in biopsy samples from the descending, ascending and transverse colon did not vary significantly  
30 between these two groups of individuals ( $p = 0.06$ ,  $0.22$  and  $0.52$  respectively). Most of the differences in rectosigmoid biopsy samples were contributed by just five of these genes (Table 1): PPAR- $\gamma$ , SAA1, IL-8, COX-2 and PPAR- $\delta$ . Similar to the alterations of gene expression in the MNM of cancer patients, we found that the expression levels of SAA1, IL-8 and COX-2 were up-regulated and those of PPAR- $\gamma$  and PPAR- $\delta$  were down-regulated  
35 in the MNM of individuals with a family history of sporadic colon cancer.

The mean ( $\pm$  SD) age in the family history group was younger ( $45 \pm 12$  years) than that of the control group ( $56 \pm 16$  years), presumably because of heightened awareness of the need for early colonoscopy in the group with a family history of colon cancer. In addition, there is a sex difference between these two groups (ten women and two men in the family history group versus nine women and seven men in the control group). However, we found that sex did not affect the level of gene expression ( $p=0.67$ ). Moreover, there was no correlation between age and the expression levels of SAA1, IL-8, COX2 and PPAR- $\gamma$  (all  $p > 0.05$ ) except for PPAR- $\delta$  0.01). Nevertheless, abnormal expression (down-regulation) of PPAR- $\delta$  increases with age. Thus comparison between younger family history group and older controls, would be biased toward finding fewer, rather than more, abnormal expressions in the family history group. In other words, we may underestimate the incidence of altered expression of PPAR- $\delta$  in the family history group.

**Table 1. Gene expression levels in normal rectosigmoid biopsy samples from individuals with family history of colorectal cancer as compared with controls**

Genes	Controls (n=104)		Patients with family history (n=77)		P Values
	Range	Mean $\pm$ (S.D.)	Range	Mean $\pm$ (S.D.)	
PPAR- $\gamma$	0.44 - 1.65	$1.07 \pm 0.41$	0.20 - 2.59	$0.79 \pm 0.40$	0.006
SAA1	0.17 - 22	$2.16 \pm 3.67$	0.33 - 2343	$151 \pm 452$	0.02
IL-8	0.14-13	$1.71 \pm 1.94$	6.84-13	$6.84 \pm 2.82$	0.02
COX-2	0.17 - 18	$1.82 \pm 2.75$	0.24 - 30	$5.11 \pm 9.01$	0.07
PPAR- $\delta$	0.39 - 2.66	$1.11 \pm 0.48$	0.16 - 2.22	$0.89 \pm 0.46$	0.07
CD44	0.35 - 4.13	$1.14 \pm 0.64$	0.11 - 4.98	$1.41 \pm 0.78$	0.12
c-Myc	0.24 - 3.66	$1.21 \pm 0.75$	0.26 - 4.31	$1.48 \pm 0.82$	0.14
MCSF-1	0.38-22	$1.81 \pm 2.59$	0.20-11	$2.04 \pm 2.19$	0.21
Gro- $\alpha$	0.01 - 51	$2.61 \pm 5.48$	0.34-57	$5.76 \pm 11.63$	0.22
Gro- $\gamma$	0.16-35	$2.18 \pm 4.29$	0.12-41	$2.55 \pm 5.91$	0.25
P21	0.51 - 2.15	$1.10 \pm 0.62$	0.20-7.68	$0.90 \pm 0.32$	0.27
PPAR- $\alpha$	0.31 - 2.38	$1.09 \pm 0.55$	0.26-2.21	$1.00 \pm 0.40$	0.54
CXCR2	0.22 - 13	$1.45 \pm 1.78$	0.43 - 4.44	$1.49 \pm 1.55$	0.55
OPN	0.19 - 13	$1.66 \pm 2.05$	0.15 - 12	$1.41 \pm 1.92$	0.73
CyclinD	0.34 - 3.48	$1.28 \pm 0.85$	0.13 - 3.21	$1.29 \pm 0.79$	0.81
COX-1	0.27 - 5.97	$1.21 \pm 0.85$	0.25 - 2.63	$1.09 \pm 0.51$	0.87

Comparison with cut-off points for "normal" gene expression

Relative gene expression levels in the rectosigmoid samples varied among individuals, much more so in samples obtained from the individuals with a family history of

colon cancer than the corresponding values from the controls (Table 1). We therefore use the expression level of each gene in the control group to define the "normal" expression level for each gene by calculating a cut-off point ( $p = 0.01$ ) for each gene. Figure 3 shows the distribution of the log (base2) expression values for genes, PPAR- $\gamma$ , IL-8, SAA 1 and COX-2 and their cut-off points. As expected, less than 1% of the biopsy samples from the control group had expression of these genes above or below the cut-off lines ( $p = 0.01$ , Figure 3). However, 21%, 12% and 8% of the biopsy samples from the family history group had expression of SAA1, IL-8 and COX-2, respectively, above the cut-off points, and 12% of them had expression of PPAR- $\gamma$  below the cut-off point (Table 2).

10

**Table 2. Number of biopsy samples (N) with gene expression above/below the cut-off point in normal individuals and individuals with a family history of colon cancer**

Genes	Biopsy samples from Normal Controls (n=104) N (%)	Biopsy samples from individuals with Family History (n= 77) N (%)
PPAR- $\gamma$	0	9 (12%)†‡
SAAI	0	16(21%)*‡
IL-8	0	9 (12%)*‡
COX-2	1 (1%)*	6 ( 8%)*‡
PPAR- $\delta$	0	2 (3%)†
Gro- $\gamma$	1 (1%)*	2 (3%)*
PPAR- $\alpha$	0	2 (3%)†
Gro- $\alpha$	0	0
MCSF-1	1 (1%)*	0
OPN	1 (1%)*	0
P21	0	0
CD44	1 (1%)*	0
CXCR2	1 (1%)*	0
c-Myc	0	0
CyclinD	0	0
COX-1	0	0

15

† with gene expression level below the cut-off point

\* with gene expression level above the cut-off point

‡ number of patients with alterations are listed in Table 3.

We next analyzed each individual in the family history group (Table 3). The number of biopsy samples which exhibited expression levels below (for PPAR- $\gamma$  and  $\delta$ ) or above (for IL-8, SAA1 and COX-2) the cut-off point ( $p=0.01$ ) are indicated. Individuals with all the biopsy samples exhibiting expression levels within the normal range are indicated with a (-) sign. All the grandparents with colon cancers in this study are maternal. Ages of the family member when colon cancer was diagnosed are indicated as follows: \*\*\* indicates that colon cancer was diagnosed before 50 years of age; \*\* indicates before 60 years of age; and \* indicates after 60 years of age. Ages of the rest of the family members when colon cancer was diagnosed are not available. None of the twelve patients in the family history group reported other types of cancer in the family except that father of the patient for case #10 had lung cancer in the 1970's.

As evidenced in Table 3, for the five most commonly altered genes, nine of the twelve individuals with a family history of colon cancer had at least one biopsy sample with expression levels below or above the cut-off point. Two individuals (cases #1 and 2) had altered expression of three of these genes in apparently normal rectosigmoid mucosa. In contrast, only one of the sixteen individuals in the control group had altered expression of one of these five genes (see Table 2). The cut-off is set so that 1% of expressions could be false positives. However, the numbers of biopsy samples obtained from each individual are different. To make an adjustment for the number of specimens, we also calculated, for each case, the probability that the number of observed samples outside the upper 99 percentile was due to chance. This calculation was based on the binomial distribution. As shown in Table 3, the observed altered gene expression in seven of the twelve individuals of the family history group is unlikely due to chance ( $p < 0.01$ ). In these seven cases, expressions of at least two of the five genes were altered. In addition, among the sixteen genes analyzed, PPAR- $\gamma$  and SAA1 are the most frequently altered genes that occurred in five of the twelve individuals with a family history of colon cancer (Table 3).

Table 3. Summary of Expression of PPAR- $\gamma$ , IL-8, SAA1, COX-2 and PPAR- $\delta$  in Rectosigmoid Biopsy Samples from Individuals with a Family History of Colon Cancer

Case	Sex	Age (years)	Family member with cancer	# of biopsy samples analyzed	# of samples with altered expression					# of genes with altered expression	Probability that changes are due to chance
					PPAR- $\gamma$	SAA1	IL-8	COX-2	PPAR- $\delta$		
1	F	53	mother***	2	2	2	-	-	1	3	<0.001
2	F	53	mother*	6	2	-	1	-	1	3	<0.001
3	M	43	father*	5	3	1	-	-	-	2	<0.001
4	F	47	mother*	7	-	7	1	-	-	2	<0.001
5	F	52	mother	8	-	-	-	-	-	0	1
6	F	52	Father and daughter***	6	-	-	1	-	-	1	0.26
7	F	18	grandfather and sister***	8	2	-	-	1	-	2	<0.01
8	F	35	Mother* and grandmother	8	-	-	8	6	-	2	<0.001
9	F	46	father**	8	-	-	-	-	-	0	1
10	F	64	sister*	6	-	1	-	-	-	1	0.26
11	F	36	mother and grandfather	7	-	-	-	-	-	0	1
12	M	38	mother and grandfather	6	1	6	-	-	-	2	<0.001

# of individuals with altered gene expression

5 5 4 2 2

Expression of different genes are altered at different sites of MNCM from individuals with a family history of colon cancer.

Analysis of individual cases from the family history group showed that different genes were altered in rectosigmoid biopsy samples in different subjects. For instance, SAA1 and PPAR- $\gamma$  were altered in case #3, IL-8 and SAA1 were altered in case #4; while COX-2 and IL-8 but not SAA1 were altered in case #8 (Figure 4A). In addition, some genes were altered in all the rectosigmoid biopsy samples from the same patient (such as SAA 1 in case #4 and IL-8 in case #8), while others were only altered in some of these biopsy samples (*i.e.* SAA1 and PPAR- $\gamma$  in case #3, IL-8 in case #4 and COX-2 in case #8). In addition, some of these alterations are restricted to the rectosigmoid regions; such as IL-8 in case #4; while others can be extended to other regions of the colon, such as SAAI in case #4 (Figure 4B).

We also observed that the difference in gene expression between the two groups of individuals increased along the length of the colon for PPAR- $\gamma$  ( $p=0.001$  for trend) and SAA1 ( $p < 0.001$ ), but not for IL-8 ( $p = 0.20$ ), COX2 ( $p = 0.58$ ), nor PPAR- $\delta$  ( $p = 0.54$ ). These results suggest that there is an increasing abnormality along the colon going from the ascending to the rectal portion between the two groups of individuals that can be detected despite reduced numbers of samples toward the ascending portion in this study.

From the foregoing example, it was possible to draw the following conclusions. Approximately 5-10% of colorectal cancers occur among patients with one of the two autosomal dominant hereditary forms of colon cancer (familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer), or who have inflammatory bowel disease (Burt R., Peterson G.M. In: Young G., Rozen, P. & Levin, B. Saunders, ed. in *Prevention and Early Detection of Colorectal Cancer*, Philadelphia, 171-194 (1996)). Of the remaining colon cancers, approximately 20% are associated with a family history of colon cancer, which is associated with a two-fold increased risk of developing colon cancer (Smith R.A., von Eschenbach A.C., Wender R., *et al.*, *American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers*, and *Update 2001--testing for early lung cancer detection*, 51 CA Cancer J Clin. 38-75; quiz 77-80 (2001)). Although linkage to chromosomes 15q13-14 and 9q22.2-31.2 has been reported in a subset of patients with familial colorectal cancer (Wiesner G.L., Daley D., Lewis S., *et al.*, *A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2*, 100 Proc Natl Acad Sci U S A, 12961-5 (2003)), the genetic basis for most of these cases is not known. In this study, we have demonstrated substantial alterations in the expression of PPAR- $\gamma$ , IL-8 and SAAI in the rectosigmoid MNCM from individuals with a family history of sporadic colon cancer, even though these individuals had no detectable colon abnormalities. Our previous study showed that, in addition to PPAR- $\gamma$ , IL-

8 and SAA1, expressions of PPAR- $\delta$ , p21, OPN, COX-2, CXCR2, MCSF-1 and CD44 were also altered significantly in the MNCM of colon cancer patients when compared to normal controls without colon cancer, polyps, or family history. These observations suggest that altered expression of genes related to cancer development in the MNCM may be a sequential event and may occur earlier than the appearance of gross morphological abnormalities. For example, altered expression of PPAR- $\gamma$ , SAA1 and IL-8 may occur in MNCM of individuals who have not developed colon cancer, but are at high risk of doing so; while altered expressions of other genes, such as PPAR- $\delta$ , p21, OPN, COX-2, CXCR2, MCSF-1 and CD44, may occur later in MNCM of individuals who have already developed a colon cancer (Chen L-C, Hao C-Y, Chiu Y.S.Y., *et al.*, *Alteration of Gene Expression in Normal Appearing Colon Mucosa of APC<sup>min</sup> Mice and Human Cancer Patients*, 64 Cancer Research 3694-3700 (2004)).

Genetic and epigenetic changes have been reported in macroscopically normal tissues for several neoplasms (Tycko B., *Genetic and epigenetic mosaicism in cancer precursor tissues*, 983 Ann N Y Acad Sci., 43-54 (2003)). For example, allelic loss has been demonstrated in normal breast terminal ductal lobular units adjacent to primary breast cancers. (Deng G., Lu Y., Zlotnikov G., Thor A.D., Smith H.S., *Loss of heterozygosity in normal tissue adjacent to breast carcinomas*, 274 Science, 2057-9 (1996)). Such allelic loss is associated with an increased risk of local recurrence (Li Z., Moore D.H., Meng Z.H., Ljung B.M., Gray J.W., Dairkee S.H., *Increased risk of local recurrence is associated with allelic loss in normal lobules of breast cancer patients*, 62 Cancer Res., 1000-3 (2002)). In addition, normal-appearing colonic mucosal cells from individuals with a prior colon cancer are more resistant to bile acid-induced apoptosis than mucosal cells from individuals with no prior colon cancer (Bernstein C., Bernstein H., Garewal H., *et al.*, *A bile acid-induced apoptosis assay for colon cancer risk and associated quality control studies*, 59 Cancer Res., 2353-7 (1999); and Bedi A., Pasricha P.J., Akhtar A.J., *et al.*, *Inhibition of apoptosis during development of colorectal cancer*, 55 Cancer Res., 1811-6 (1995)). Since apoptosis is important in colonic epithelium to eliminate cells with unrepaired DNA damage (Payne C.M., Bernstein H., Bernstein C., Garewal H., *Role of apoptosis in biology and pathology: resistance to apoptosis in colon carcinogenesis*, 19 Ultrastruct Pathol., 221-48 (1995)), reduction in apoptosis could result in the retention of DNA-damaged cells and increase the risk of carcinogenic mutations.

PPAR- $\gamma$  is down-regulated in several carcinomas. Ligands of PPAR- $\gamma$  inhibit cell growth and induce cell differentiation (Kitamura S., Miyazaki Y., Shinomura Y., Kondo S., Kanayama S., Matsuzawa Y., *Peroxisome proliferator-activated receptor gamma induces growth arrest and differentiation markers of human colon cancer cells*, 90 Jpn J Cancer Res 75-80 (1999)), and loss-of-function mutations in PPAR- $\gamma$  have been reported in human colon cancer (Sarraf P., Mueller E., Smith W.M., *et al.*, *Loss-of-function mutations in PPAR gamma*

associated with human colon cancer, 3 Mol. Cell, 799-804 (1999)). Thus, our observation of down-regulation in PPAR- $\gamma$  expression in MNCM may represent an early event that promotes colonic epithelial cell growth and inhibits cellular differentiation. In addition, PPAR- $\gamma$  also negatively regulates inflammatory response (Welch J.S., Ricote M., Akiyama T.E., Gonzalez F.J., Glass C.K., *PPAR gamma and PPAR delta negatively regulate specific subsets of lipopolysaccharide and IFN-gamma target genes in macrophages*, 100 Proc Natl Acad Sci U S A 6712-7 (2003)). Inflammation favors tumorigenesis by stimulating angiogenesis and cell proliferation (Nakajima N., Kuwayama H., Ito Y., Iwasaki A., Arakawa Y., *Helicobacter pylori*, neutrophils, interleukins, and gastric epithelial proliferation, 25 Suppl. 1 J Clin Gastroenterol., 98-202 (1997)). Similarly, IL-8 and the acute-phase protein SAA1 modulate the inflammatory process (Dhawan P., Richmond A., *Role of CXCL 1 in tumorigenesis of melanoma*, 72 J Leukoc Biol., 9-18 (2002); and Urieli-Shoval S., Linke R.P., Matzner Y., *Expression and function of serum amyloid A, a major acute-phase protein, in normal and disease states*, 7 Curr Opin Hematol., 64-9 (2000)). Up-regulation of pro-inflammatory cytokines and acute phase proteins has been reported in the colon mucosa of individuals with inflammatory bowel disease (Niederau C., Backmerhoff F., Schumacher B., *Inflammatory mediators and acute phase proteins in patients with Crohn's disease and ulcerative colitis*, 44 Hepatogastroenterology, 90-107 (1997); and Keshavarzian A., Fusunyan R.D., Jacyno M., Winship D., MacDermott R.P., Sanderson I.R., *Increased interleukin-8 (IL-8) in rectal dialysate from patients with ulcerative colitis: evidence for a biological role for IL-8 in inflammation of the colon*, 94 Am J Gastroenterol., 704-12 (1999)), who are at very high risk of developing colon cancer (Bachwich D.R., Lichtenstein G.R., Traber P.G., *Cancer in inflammatory bowel disease*, 78 Med Clin North Am., 1399-412 (1994)). Epidemiological observations also suggest that chronic inflammation predisposes to colorectal cancer (Rhodes J.M., Campbell B.J., *Inflammation and colorectal cancer: IBD-associated and sporadic cancer compared*, 8 Trends Mol Med., 10-6 (2002); and Farrell R.J., Peppercorn M.A., *Ulcerative colitis*, 359 Lancet 331-40 (2002)). Thus, the observation of down-regulation of PPAR- $\gamma$  and up-regulation of IL-8 and SAA1 in the normal mucosa of individuals with a family history of sporadic colon cancer and individuals with inflammatory bowel disease may indicate the involvement of common pathways leading to colon carcinogenesis in these two groups.

Our observation of altered expression of genes associated with cancer and inflammation in normal colonic mucosa in some individuals with a family history of colon cancer is consistent with the recent report of association of elevated serum C-reactive protein ("CRP") concentration prior to the development of colon cancer (Erlinger T.P., Platz E.A., Rifai N., Helzlsouer K.J., *C-reactive protein and the risk of incident colorectal cancer*, 291 JAMA, 585-90 (2004)). These findings suggest that inflammation is a risk factor for the development

-21-

of colon cancer in average-risk individuals (*id.*). However, CRP is a nonspecific marker of inflammation that may indicate inflammation in tissues other than colon. In our study, we have analyzed the tissue where colon cancer arises and would be more specific in assessing the risk of developing colon cancer.

5 We do not know which cell type is responsible for the observed altered gene expression. There are many cell types in the colonic mucosa, including several types of mucosal epithelial cells, stromal cells and blood-born cells. Studies from our group and others have demonstrated that the up-regulation of COX-2 protein in MNCM is localized primarily to the infiltrating macrophages and secondarily to the epithelial cells in aberrant crypt foci in the MNCM of APC<sup>min</sup> mice (Chen L-C, Hao C-Y, Chiu Y.S.Y., *et al.*, *Alteration of Gene Expression in Normal Appearing Colon Mucosa of APC<sup>min</sup> Mice and Human Cancer Patients*, 64 Cancer Research 3694-3700 (2004); and Hull M.A., Booth J.K., Tisbury A., *et al.*, *Cyclooxygenase 2 is up-regulated and localized to macrophages in the intestine of Min mice*, 79 Br J Cancer, 1399-405 (1999)). From our previous studies of MNCM of APC<sup>min</sup> mice, 10 detection of the gene products that are up- or down- regulated in MNCM by immunohistochemical staining was found to be technically difficult, perhaps because the secreted proteins, such as IL-8 and SAA1, are evanescent in tissue sections (Chen L-C, Hao C-Y, Chiu Y.S.Y., *et al.*, *Alteration of Gene Expression in Normal Appearing Colon Mucosa of APC<sup>min</sup> Mice and Human Cancer Patients*, 64 Cancer Research 3694-3700 (2004)). Due to 15 the limited amount of the biopsy samples and technical difficulties, we were unable to perform immunohistochemical staining to demonstrate the cell types contributing to the altered gene expression. If the absolute RNA quantities are sufficient, RNA *in situ* hybridization may be a better method to determine the cellular locations of alterations. Alternatively, laser microdissection followed by RT-PCR may be able to define the cell types involved. 20 Regardless of the cell types responsible for the altered gene expression, our results demonstrate that relative to normal individuals without family history of colon cancer, altered gene expression is present in normal colon mucosa of some individuals with a family history of colon cancer and these individuals are known to have an increased risk of developing colon cancer (Burt R., Peterson G.M. In: Young G., Rozen, P. & Levin, B. Saunders, ed. in 25 *Prevention and Early Detection of Colorectal Cancer*, Philadelphia, 171-194 (1996)). 30

Among patients with altered gene expression in the rectosigmoid biopsy samples, some showed alterations in all biopsy samples (*i.e.*, expression of SAA1 in cases #4 and 12), while others showed altered expression in some biopsy samples only (*i.e.*, PPAR- $\gamma$  in cases #2 and #3, figure 2). Since most samples were assayed with multiple genes in duplications to 35 ensure the quality of cDNA, such heterogeneity is unlikely due to technical variation. We speculate that this heterogeneity might reflect the frequency and/or the distribution of "hot

-22-

spots" in these individuals. It is possible that the individuals with altered gene expression in all rectosigmoid biopsy samples may have wide-spread molecular abnormalities in their rectosigmoid mucosa, while those with altered expression in some of the biopsy samples have discrete hot spots. Thus, individuals in the former group may have a global predisposition to development of colon polyps or cancer, while those in the latter group may have local predisposition. Whether the risks in developing colon cancer or polyps differ between these two groups is unknown. In addition, altered expression of different combination of genes were observed in the rectosigmoid biopsy samples of individuals in the family history group. This observation suggests that different molecular pathways may be involved in the early stages of colon carcinogenesis. Whether altered gene expression in certain molecular pathways is associated with higher risk of polyps or cancer also remains to be determined.

Consistent with the reports of more aberrant crypt foci (the preneoplastic colonic lesions) in the distal colon than in the proximal colon of the sporadic colon cancer patients and the carcinogen-treated mice (Shpitz B., Bomstein Y., Mekori Y., *et al.*, *Aberrant crypt foci in human colons: distribution and histomorphologic characteristics*, 29 Hum Pathol., 469-75 (1998); and Salim E.I., Wanibuchi H., Morimura K., *et al.*, *Induction of tumors in the colon and liver of the immunodeficient (SCID) mouse by 2-amino-3-methylimidazo[4,5-f]quinoline (IQ)-modulation by long chain fatty acids*, 23 Carcinogenesis, 1519-29 (2002)), we found that most of the alterations in gene expression were found in the distal colon of the individuals from the family history group. We speculate that the distal colon mucosa of the susceptible individuals may be exposed to higher concentration of exogenous substances present in the stool than mucosa in other colon regions after most of the water is re-absorbed at the end of the large intestine, and such exposure may lead to higher rate of altered gene expression at this region.

We have shown that family history of colon cancer, but not age or sex, is the factor responsible for the observed differences in gene expression in the rectosigmoid mucosa of the two groups. The available information did not indicate any specific difference in diet or medication between these two groups of patients. However, we cannot eliminate the possibility that diet or medication affect gene expression without further study. Not all individuals with a family history of colon cancer will develop cancer or adenomatous polyps of the colon (Smith, R.A., von Eschenbach A.C., Wender, R., *et al.*, *American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers, and Update 2001—testing for early lung cancer detection*, 51 CA Cancer J. Clin., 38-75; quiz 77-80 (2001).). Consistent with this clinical observation, our analysis also showed that not all the individuals with a family history of colon

-23-

cancer have altered gene expression in MNCM. Since the genes analyzed in this study are involved in the development of colon cancer, we hypothesize that individuals with altered gene expression in the MNCM may be more susceptible to developing polyps or cancer than those without altered gene expression. To test this hypothesis, a prospective study with a larger number of study subjects will be needed. If such an association is confirmed, it may be possible to identify individuals at increased risk of developing colon cancer by using gene expression analysis of rectosigmoid biopsy samples. Theoretically, it is easier to identify individuals with global alterations in the MNCM than individuals with local alterations by analysis of random MNCM samples. However, if an appropriate panel of genes was selected for analysis using multiple samples, it may have enough predictive power to identify such patients.

Turning now to **Fig. 5**, various aspects of **Fig. 5** may be implemented using a conventional general purpose or specialized digital computer(s) and/or processor(s) programmed according to the teachings of the present disclosure, as will be apparent to those skilled in the computer arts. Appropriate software coding can be prepared readily by skilled programmers based on the teachings of the present disclosure, as will be apparent to those skilled in the software arts. The invention also may be implemented by the preparation of integrated circuits and/or by interconnecting an appropriate network of component circuits, as will be readily apparent to those skilled in the arts.

Various aspects include a computer program product which is a storage medium having instructions and/or information stored thereon/in which can be used to program a general purpose or specialized computing processor(s)/device(s) to perform any of the features presented herein. The storage medium can include, but is not limited to, one or more of the following: any type of physical media including floppy disks, optical discs, DVDs, CD-ROMs, microdrives, magneto-optical disks, holographic storage devices, ROMs, RAMs, EPROMs, EEPROMs, DRAMs, PRAMS, VRAMs, flash memory devices, magnetic or optical cards, nano-systems (including molecular memory ICs); paper or paper-based media; and any type of media or device suitable for storing instructions and/or information. Various aspects include a computer program product that can be transmitted in whole or in parts and over one or more public and/or private networks wherein the transmission includes instructions and/or information which can be used by one or more processors to perform any of the features presented herein. In various aspects, the transmission may include a plurality of separate transmissions.

Stored on one or more of the computer readable medium (media), the present disclosure includes software for controlling both the hardware of general purpose/specialized computer(s) and/or processor(s), and for enabling the computer(s) and/or processor(s) to

interact with a human user or other mechanism utilizing the results of the present invention. Such software may include, but is not limited to, device drivers, operating systems, execution environments/containers, user interfaces and applications.

The execution of code can be direct or indirect. The code can include compiled, interpreted and other types of languages. Unless otherwise limited by claim language, the execution and/or transmission of code and/or code segments for a function can include invocations or calls to other software or devices, local or remote, to do the function. The invocations or calls can include invocations or calls to library modules, device drivers and remote software to do the function. The invocations or calls can include invocations or calls in distributed and client/server systems.

**Fig. 6** depicts an aspect of this disclosure having a swab sampling and transport system **400** for the minimally invasive sampling of colonic mucosal cells. The system **400** of **Fig. 6** is comprised of a swab **410** and a container **420**. A container **420**, such as one depicted by the aspect of the disclosure shown in **Fig. 6**, is configured to stabilize, extract, and store the sample of colonic mucosal cells until the diagnostic test for early detection of CRC using the disclosed biomarker panel can be done on the sample.

The swab **410** has a tip **412** extending from the end of a shaft **414**. The tip **410** may be of a number of shapes such as oblate, square, rectangular, round, etc., and has a maximum width of about 0.5 cm to 1.0 cm, and a length of about 1.0 cm to 10.0 cm around the end of the rod. The tip **412** may be composed of a number of materials, such as cotton, rayon, polyester, and polymer foam, for example, or combinations of such materials. The shaft **414** is made of a material with sufficient mechanical strength for effectively swabbing the rectal area, but with enough flexibility to prevent injury. Examples of shaft materials having the strength and flexibility properties for a rectal swab include wood, paper, and a variety of polymeric materials, such as polyester, polystyrene, and polyurethane, and composites of such polymers.

The container **420** has a body **412** and a cap **424**. The body **412** may have a variety of lengths and diameters to accommodate a swab **410** having dimensions of the tip **412** and the range of lengths of the shaft **414** as described in the above. The body **412** of the container may be made of a number of polymeric materials, such as polyethylene, polypropylene, polycarbonate, polyfluorocarbon, or glass, while the cap **424** typically is made of a desirable polymeric material, such as the examples given for the body **412**. The container **420** has a reagent **426** in the bottom that is suitable for stabilizing and extracting the colonic mucosal cells collected on the swab **410** when swabbing of the rectal area is done as a minimally invasive sampling technique. Additionally, a container **420** having a reagent **426**

-25-

suitable for stabilizing and extracting a sample of colonic mucosal cells from a stool sample may be used without the need for the swab **410**.

The reagent **426** contains a buffered solution of guanidine thiocyanate in a concentration of at least about 0.4M and other tissue denaturing reagents such as a biological surfactant in a concentration of at about between 0.1 to 10%. Desirable biological surfactants can be zwitterionic, such as CHAPS or CHAPSO, non-ionic, such as TWEEN, or any of the alkylglucoside surfactants, or ionic, such as SDS. A variety of buffers, for example, those generally known as Good's buffers, such as Tris, may be used. The concentration of the buffer may vary in order to buffer the reagent **426** effectively to a pH of between about 7.0 to 8.5.

It is further contemplated that the sample taken using an aspect of the disclosure as in **Fig. 6** of a swab sampling and transport system **400** can be processed and the data analyzed in a single apparatus using the computer hardware and software disclosed above. That is, the sample obtained from the aspect of the disclosure of **Fig. 6** can be analyzed according to **Fig. 5** in a single apparatus. However, it is also contemplated that a patient's blood or stool sample can be analyzed in the single apparatus. In one embodiment, one aspect of the apparatus is a first component that is used to carry out RT-PCR for a sample from a patient for gene expression profiling, as described above. Gene expression profiling allows quantifying of cDNA of SEQ. ID Nos 1-16, which is reverse-transcribed from mRNA made by cells in the sample from the patient. The sets of primers from SEQ. ID Nos 33-64 are used in the RT-PCR reaction to prime strands of mRNA corresponding to SEQ. ID Nos 1-16, and thereby to synthesize cDNA corresponding to SEQ. ID Nos 1-16.

After obtaining the cDNAs from the RT-PCR, data are compared by a second component of the apparatus to control data already stored in the apparatus on a storage medium. Multivariate analysis as disclosed above is applied using software to execute instructions for the ANOVA, M-Dist, or other means of multivariate analysis. Based on the statistical analysis, a qualified diagnostician can assess the presence or absence of CRC, the progress of CRC, and/or the effects of treatment of CRC.

In a further aspect of this disclosure, protein expression profiling of patient samples can be carried out for early detection of CRC, using a single apparatus. The term "polypeptide" or "polypeptides" is used interchangeably herein with the term "protein" or "proteins." As discussed previously, proteins long have been investigated for their potential as biomarkers, with limited success. There is value in protein biomarkers as complementary to polynucleotide biomarkers. Reasons for having the information provided by both types of biomarkers include the current observations that mRNA expression levels are not good predictors of protein expression levels, and that mRNA expression levels tell nothing of the

post-translational modifications of proteins that are key to their biological activity. Therefore, in order to understand the expression levels of proteins, and their complete structure, the direct analysis of proteins is desirable.

Disclosed herein are proteins listed in SEQ. ID NOs 17-32, which correspond to the  
5 genes indicated in SEQ. ID NOs 1-16. A further aspect of the disclosed invention is to determine expression levels of the proteins indicated by SEQ. ID NOs. 17-32. A sample from the patient, taken by non- or minimally-invasive methods as disclosed above, can be used to prepare fixed cells or a protein extract of cells from the sample. The cells for protein expression profiling can be obtained either through the method of **Fig. 6**, or alternatively for  
10 example by a blood sample or stool sample, or other non-invasive or minimally invasive method (or of course by more conventional invasive methods, including for example sigmoidoscopy and other procedures).

In a first component of the apparatus, the cells or protein extract can be assayed with a panel of antibodies – either monoclonal or polyclonal – against the claimed panel of  
15 biomarkers for measuring targeted polypeptide levels. The objective of the assay is to detect and quantify expression of proteins corresponding to the biomarker gene sequences in SEQ. ID NOs 1-16, *i.e.*, SEQ. ID NOs 17-32.

In one aspect of the disclosure contemplated for the method, the antibodies in the antibody panel, which are based on the panel of biomarkers, can be bound to a solid support.  
20 The method for protein expression profiling may use a second antibody having specificity to some portion of the bound, targeted polypeptide. Such second antibody may be labeled with molecules useful for detecting and quantifying the bound polypeptides, and therefore in binding to the polypeptide, label it for detection and quantification. Additionally, other reagents are contemplated for labeling the bound polypeptides for detection and  
25 quantification. Such reagents may either directly label the bound polypeptide or, analogous to a second antibody, may be a moiety with specificity for the bound polypeptide having labels. Examples of such moieties include but are not limited to small molecules such as cofactors, substrates, complexing agents, and the like, or large molecules such as lectins, peptides, oligonucleotides, and the like. Such moieties may be either naturally occurring or synthetic.

30 Examples of detection modes contemplated for the disclosed methods include, but are not limited to spectroscopic techniques, such as fluorescence and UV-Vis spectroscopy, scintillation counting, and mass spectroscopy. Complementary to these modes of detection, examples of labels for the purpose of detection and quantitation used in these methods include, but are not limited to chromophoric labels, scintillation labels, and mass labels. The  
35 expression levels of polynucleotides and polypeptides measured in a second component of the apparatus using these methods may be normalized to a control established for the

-27-

purpose of the targeted determination. The control data is stored in a computer which is a third component of the apparatus.

A fourth software component compares the data obtained from a patient's or a plurality of patients' samples to the control data. The comparison will comprise at least one  
5 multivariate analysis, and can include ANOVA, MANOVA, M-Dist, and others known to those of ordinary skill in the art. Once the statistical analysis and comparison is performed and complete, a physician or other qualified person can make a diagnosis concerning the patient's or patients' CRC status.

Turning now to the drug screening aspect of the present disclosure, it is noted that the  
10 panel of biomarkers disclosed herein are genes and expression products thereof that also are known to be involved in the following metabolic pathways and processes: 1) oxidative stress/inflammation; 2) APC/b-catenin pathway; 3) cell cycle/ transcription factors; and 4) actions of cytokines and other factors involved in cell/cell communications, growth, repair and response to injury or trauma. There is increasing evidence that these pathways, and hence  
15 members of the subject panel of biomarkers, are also involved in many other kinds of cancers than CRC, such as lung, prostate and breast, as well as neurodegenerative diseases, such as Alzheimer's and amyotrophic lateral sclerosis ("ALS"). In such pathologies, genes and expression products thereof involved in these pathways are fundamental to the growth, maintenance and response to stress of cells of many different types. During a pathology such  
20 as cancer or neurodegeneration, altered expression of certain altered genes results in a pathological symptom or symptoms, so that a shift in those genes, and expression products thereof, are characteristic biomarkers of that particular pathology. In that regard, seemingly unrelated pathologies, such as various cancers and neurodegenerative diseases, are manifestations of very complex pathologies that each involve discrete members of the subject  
25 biomarkers, which are genes and expression products thereof drawn from the above group of pathway and processes. As practical evidence of this, it is now appreciated that COX-2 inhibitors have therapeutic value for a wide variety of disorders, including not only colon and other cancers, but for some neurodegenerative diseases as well.

What is disclosed herein is the use of the subject biomarker panel in **Fig. 1** in the drug  
30 discovery process for pathologies such as cancers, for example CRC, lung prostate, and breast, and neurodegenerative diseases, for example Alzheimer's and ALS. As mentioned in the above, the discrete pattern of altered genes and expression products thereof provides a unique signature for each specific disease, so the panel provides the necessary selectivity for a variety of pathologies. What is meant by drug is any therapeutic agent that is useful in the  
35 treatment of a pathology. This includes traditional synthetic molecules, natural products,

-28-

natural products that are synthetically modified, and biopharmaceutical products, such as polypeptides and polynucleotides, and combinations, extracts and preparations thereof.

Drug screening is part of the first stage of drug development referred to as the drug discovery phase. Prospective drugs that are qualified through the drug screening process are typically referred to as leads, which is to say that in passing the criteria of the screening process they are advanced to further testing in a stage of drug discovery generally referred to as lead optimization. If passing the lead optimization stage of drug discovery, the leads are qualified as candidates, and are advanced beyond the drug discovery stage to the next stage of drug development known as preclinical trials, and are referred to as investigative new drugs ("IND"). If the IND is advanced, it is advanced to clinical trials, where it is tested in human subjects. Finally, if the IND shows promise through the clinical trial stage, after approval from FDA, it may be commercialized. The entire drug development process for a single candidate is known to take 10-15 years and hundreds of millions of dollars in development costs. For that reason, the current strategy within the pharmaceutical drug development community is to focus on the drug discovery stage as effective in weeding out prospective drugs efficiently, and advancing only candidates with high potential for success through the remaining drug development cycle.

In the screening stage of drug discovery, a specific assay for evaluating prospective drugs is performed against a qualified biological model system for which a specific endpoint is monitored. A biomarker panel that is used as a surrogate endpoint for drug screening for pathologies, such as cancers, for example CRC, lung, prostate, and breast, and neurodegenerative diseases, for example Alzheimer's and ALS, is not only a panel useful for early detection of such pathologies, but additionally demonstrates modulation by a drug in a fashion that correlates with a decrease in the pathology occurrence or recurrence. Additionally, one or more members of a biomarker panel useful in the early detection of such pathologies may also be useful as targets for drug screening for such pathologies. As will be discussed subsequently, the biomarkers described by **Fig. 1** may be useful both as surrogate endpoints in model biological systems, as well as targets in drug screening.

During the screening phase, large libraries of prospective drugs may be evaluated, representing a throughput of tens of thousands of compounds over a single screening regimen. What is regarded as low-throughput screening ("LTS") is about 10,000 to about 50,000 prospective drugs, while medium-throughput screening ("MTS") represents about 50,00 to about 100,00 prospective drugs, and high-throughput screening ("HTS") is 100,000 to about 500,000 prospective drugs.

What is meant by screening regimen includes both the testing protocol and analytical methodology by which the screening is conducted. The screening regimen, then, includes

-29-

factors such as the type of biological model that will be used in the test; the conditions under which the testing will be conducted; the type of prospective drug candidates, or library of prospective candidates that will be used; the type of equipment that will be used; and the manner in which the data are collected, processed, and stored. The scale of the screening regimen --LTS, MTS, or HIS -- is impacted by factors such as testing protocol (e.g., type of assay), analytical methodology (e.g., miniaturization, automation), and computational capability and capacity. What is meant by biological model system includes whole organism, whole cell, cell lysate, and molecular target. What is meant by prospective drug candidate is any type of molecule, or preparation or suspension of molecules, under consideration for having therapeutic use. For example, the prospective drug candidates could be synthetic molecules, natural products, natural products that are synthetically modified, and biopharmaceutical products, such as polypeptides and polynucleotides, and combinations, extracts, and preparations thereof.

As discussed above, **Fig. 1** provides sequence listings of a panel of biomarkers useful in practicing the disclosed invention. One aspect of the disclosure is a biomarker panel of 16 identified coding sequences given in SEQ. ID NOs 1-16, while another aspect of a biomarker panel is the 16 identified proteins given by SEQ. ID NOs 17-31. These two aspects of the present invention provide the selectivity and sensitivity necessary for the early detection of pathologies, such as cancers, for example CRC, lung, prostate, and breast, and neurodegenerative diseases, for example Alzheimer's and ALS.

As previously mentioned, CRC is an exemplary pathology contemplated for development of novel drugs. For CRC, no biomarker or biomarker panel has been identified that has an acceptably high degree of selectivity and sensitivity to be effective for early detection of CRC. Therefore, what is described in **Fig. 1** are aspects of biomarker panels that are differentiating in providing the basis for early detection of CRC. Selectivity of a biomarker defined clinically refers to percentage of patients correctly diagnosed. Sensitivity of a biomarker in a clinical context is defined as the probability that the disease is detected at a curable stage. Ideally, biomarkers would have 100% clinical selectivity and 100% clinical sensitivity. To date, no biomarker or biomarker panel has been identified that has an acceptably high degree of selectivity and sensitivity required to be effective for the broad range of needs in patient care management.

The analytical methodology by which the screening is conducted may include the methodologies disclosed above for early detection of CRC, *i.e.* gene expression profiling from the mRNA of a biological sample to determine the gene expression of biomarkers and how their expression level(s) might have been affected by a prospective drug candidate (including use of RT-PCR), and/or determining protein expression levels of the **Fig. 1** polypeptide

-30-

biomarkers due to application of a prospective drug candidate; and then applying multivariate statistical analysis to determine the statistical significance of the expression levels of the various markers in the panel, with and without the prospective drug candidate(s).

Referring to **Fig. 7**, one aspect of the drug screening disclosure contemplates obtaining a tissue sample, such as a swab (see **Fig. 6**), blood sample, or biopsy, which can be taken by, for example, minimally invasive, invasive, or non-invasive means. An appropriate lysis buffer can be used to extract and preserve the RNA of the cells in the tissue sample. RT-PCR then can be carried out on the extracted RNA and converted to cDNA, as disclosed above, using, for example, at least two of the primers listed in SEQ. ID NOs 33-64, specific to the biomarker panel of **Fig. 1**, to screen the effect of the drug. The results of the assay can then be subjected to a multivariate analysis and M-dist, as disclosed above, and the results compared to control data.

**Figure 8** depicts a further aspect of the drug screening disclosure in which antibodies are made against at least two biomarker proteins listed as SEQ. ID NOs 17-32, and the antibodies are used to assay a biological system, for example whole cells, cell lysates, etc. from, for example, biopsies or other tissue samples as set forth above. The antibodies are used to detect and quantify expression of the biomarker peptides identified by SEQ. ID NOs 17-32, so that the expression of these biomarker peptides can be monitored as a function of dosing the biological system with a potential drug. The results can be subjected to multivariate or univariate analysis and M-dist., as disclosed above, and compared to control data.

What has been disclosed herein has been provided for the purposes of illustration and description. It is not intended to be exhaustive or to limit what is disclosed to the precise forms described. Many modifications and variations will be apparent to the practitioner skilled in the art. What is disclosed was chosen and described in order to best explain the principles and practical application of the disclosed embodiments of the art described, thereby enabling others skilled in the art to understand the various embodiments and various modifications that are suited to the particular use contemplated.

The references cited above are incorporated by reference in full.

-31-

**CLAIMS**

What is claimed:

- 5           1. A method for making a reagent composition for the early detection of colorectal cancer, lung cancer, prostate cancer, breast cancer, Alzheimer's and ALS, the method comprising:  
  
              synthesizing a pair of primers for each polynucleotide pair from SEQ. ID NOs 33-64;  
10               adjusting to at least one desired concentration in a plurality of separate stock solutions each of said primers, using a diluent;  
  
              aliquoting each of said stock solutions of each of said primers into a plurality of  
15               containers; and  
  
              storing the plurality of containers in long-term storage conditions.
- 20           2. The method of claim 1 wherein the method further comprises lyophilizing the aliquoted stock solutions of each of said primer pairs.
- 25           3. A method for early detection of colorectal cancer, lung cancer, prostate cancer, breast cancer, Alzheimer's and ALS, the method comprising:  
  
              obtaining a tissue sample by a non-invasive or a minimally invasive method from grossly-normal appearing tissue;  
  
              isolating RNA from the sample;  
  
30               amplifying copies of cDNA from the RNA sample using a plurality of pairs of primers selected from the group consisting of SEQ. ID NOs 33-64, to detect a panel of polynucleotides selected from SEQ. ID NOs. 1-16;  
  
              quantifying the amplified copies of cDNA; and  
35

-32-

using the quantified amplified copies of cDNA to assess at least one of disease progress and treatment effectiveness for at least one of colorectal cancer, lung cancer, prostate cancer, breast cancer, Alzheimer's and ALS.

- 5           4. The method as in claim 3 wherein the obtaining step further comprises sampling rectal mucosal cells.
5. The method of claim 3 wherein the obtaining step further comprises one of drawing blood, sampling stool, and taking a rectal biopsy.
- 10          6. The method of claim 3 wherein the using step further comprises:
- analyzing by multivariate analysis the quantified levels of tissue sample cDNA;
- 15          comparing the multivariate analysis of the quantified levels of tissue sample cDNA with a plurality of control data, wherein the comparison determines a significance of differences from the control data to assess the presence of colorectal cancer.
- 20          7. The method of claim 6 wherein the analyzing step further comprises using one of an ANOVA test and a Mahalanobis distance test.
8. A method for early detection of colorectal cancer and for evaluation of treatment efficacy of colorectal cancer, the method comprising the steps of:
- 25          obtaining by a non-invasive or minimally-invasive method a tissue sample containing cells that grossly appear cancer-free;
- generating a plurality of antibodies having different specificities against each of the polypeptides identified by SEQ. ID NOs 17-32;
- 30          assaying for expression of polypeptides in a panel of polypeptides identified by SEQ. ID NOs 17-32 with the plurality of antibodies, wherein the assaying step allows for quantifying specific binding of the antibodies to the polypeptides;
- 35          quantifying the levels of each of the different polypeptides in the panel of polypeptides based on the quantified specific antibody binding; and

-33-

analyzing the quantified levels of each of the different polypeptides in the panel of polypeptides, wherein the quantified levels are used to assess at least one of the presence, progress, and treatment of colorectal cancer.

5

9. The method of claim 8 wherein the obtaining step further comprises one of sampling blood, sampling stool, swabbing for colonic cells, and performing a rectal biopsy.

10

10. A method for analyzing data for the early detection and treatment monitoring of colorectal cancer, the method comprising the following steps:

obtaining a plurality of quantified levels of cDNA for polynucleotides selected from SEQ. ID Nos. 1-16 from a patient sample, wherein the sample is taken by a non-invasive method or a minimally-invasive method;

15

comparing said data from the patient sample to a plurality of stored control data using multivariate statistical analysis; and

20

making a determination concerning one of diagnosis of colorectal cancer, colorectal cancer progress, and treatment efficacy for the patient based on the comparison.

11. A machine readable medium having instructions stored thereon that, when executed by one or more processors, cause a system to:

25

obtain the data of quantified levels of cDNA for polynucleotides listed in SEQ. ID NOs. 1-16, wherein the quantified levels of cDNA are from a patient tissue sample and a control tissue sample;

30

compare the quantified levels of cDNA from the patient tissue sample to the quantified levels of cDNA from the control tissue sample using at least one multivariate statistical analysis; and

provide said multivariate statistical analysis for evaluation by an individual trained to evaluate colorectal cancer.

35

-34-

12. A computer signal embodied in a transmission medium, comprising:

5 a code segment including instruction for obtaining quantified levels of cDNA for polynucleotides selected from SEQ. ID NOs. 1-16, wherein the quantified levels of cDNA are from a patient tissue sample;

10 a code segment including instruction for comparing the quantified levels of cDNA from the patient tissue sample to a plurality of control data using multivariate statistical analysis; and

a code segment including instruction for making a diagnosis of colorectal cancer for the patient tissue sample based on the comparison.

13. A computer signal embodied in a transmission medium, comprising:

15 a code segment including instruction for obtaining quantified levels of polypeptides selected from SEQ. ID NOs. 17-33, wherein the quantified levels of polypeptides are from a patient sample containing colonic mucosal cells;

20 a code segment including instruction for comparing the quantified levels of polypeptides from the patient sample to a plurality of control data using multivariate statistical analysis; and

25 a code segment including at least one instruction based on the comparison for at least one of a diagnosis of colorectal cancer, a progress of colorectal cancer, and an efficacy of treatment of colorectal cancer.

14. A kit for use in the early detection of colorectal cancer, the kit comprising:

30 a collection container for receiving a sample containing rectal mucosal cells obtained through a non-invasive procedure, wherein the collection container is configured to stabilize and store the sample; and

35 at least one reagent that is used in the analysis of polynucleotide expression levels, wherein the polynucleotides are selected from SEQ. ID Nos. 1-16.

-35-

15. A kit for use in the detection of colorectal cancer, the kit comprising:

a swab sampling and sample transport system for the minimally invasive sampling of rectal mucosal cells, which system is comprised of:

5

a swab configured to sample colonic mucosal cells from the rectum; and

a collection container for receiving the swab after the sample has been taken, wherein the collection container is configured to stabilize, extract and store the sample; and

10

at least one reagent that is used in the analysis of polynucleotide expression levels, wherein the polynucleotides are selected from SEQ. ID Nos. 1-16.

15

16. A method for drug screening, the method comprising the following steps:

selecting a model biological system for at least one of colorectal cancer, lung cancer, prostate cancer, breast cancers, Alzheimer's and ALS;

20

selecting at least one prospective drug for screening using the suitable model biological system;

selecting at least two biomarkers from a panel of biomarkers identified by SEQ. ID 1-32;

25

dosing the model biological system with the at least one prospective drug; and

monitoring the response of the at least two biomarkers in the model biological system as a function of the dosing step.

30

17. The method of claim 16, further comprising: determining the efficacy of the prospective drug based on the monitoring step.

1/10

Sequence ID No. / ID	NCBI Entrez Database	Name	Abbreviation
1. Coding sequence	XM_031289	Interleukin 8	IL8
2. Coding sequence	XM_051900	Prostaglandin-endoperoxide synthase 2	PTGS2
3. Coding sequence	M94582	Interleukin 8 receptor B	ILR8RB
4. Coding sequence	NM_000757	Macrophage colony stimulating factor 1	CSF1 (MCSF1)
5. Coding sequence	X54489	Melanoma growth stimulatory activity	MGSA
6. Coding sequence	NM_002090	Chemokine (C-X-C motif) ligand 3	CXCL3
7. Coding sequence	XM_032429	Secreted phosphoprotein 1	SPP1 (OPN)
8. Coding sequence	M64349	Cyclin D	CCND1
9. Coding sequence	AX057136	c-Myc	c-Myc
10. Coding sequence	L25610	Cyclin-dependent kinase inhibitor	HUMCDK1
11. Coding sequence	NM_005036	Peroxisome proliferative activated receptor, alpha	PPARA
12. Coding sequence	XM_003059	Peroxisome proliferative activated receptor, gamma	PPARG
13. Coding sequence	NM_006238	Peroxisome proliferative activated receptor, delta	PPARD
14. Coding sequence	XM_030326	CD44 antigen	CD44
15. Coding sequence	XM_044882	Prostaglandin-endoperoxide synthase 1	PTGS1
16. Coding sequence	NM_000331	Serum amyloid A1	SAA1
17. Protein	XP_031289	Interleukin 8	IL8
18. Protein	XP_051900	Prostaglandin-endoperoxide synthase 2	COX2
19. Protein	AAA36108	Interleukin 8 receptor B	CXCR2
20. Protein	NP_000757	Macrophage colony stimulating factor 1	MCSF1
21. Protein	CAA38361	Melanoma growth stimulatory activity	Groα
22. Protein	NM-002090	Chemokine (C-X-C motif) ligand 3	Groγ
23. Protein	XP_032429	Osteopontin	OPN
24. Protein	AAA52136	Cyclin D	cyclin D1
25. Protein	CAC22425	c-Myc	c-Myc

FIG. 1

2/10

Sequence ID No. / ID	NCBI Entrez Database	Name	Abbreviation
26. Protein	AAA16109	Cyclin-dependent kinase inhibitor	p21
27. Protein	NP_005027	Peroxisome proliferative activated receptor, alpha	PPAR $\alpha$
28. Protein	XP_003059	Peroxisome proliferative activated receptor, gamma	PPAR $\gamma$
29. Protein	NP_006229	Peroxisome proliferative activated receptor, delta	PPAR $\delta$
30. Protein	XP_030326	CD44 antigen	CD44
31. Protein	XP_044882	Prostaglandin-endoperoxide synthase 1	COX1
32. Protein	NP_000331	Serum amyloid A1	SAA1
33. Forward primer		Interleukin 8	IL8
34. Reverse primer			
35. Forward primer		Prostaglandin-endoperoxide synthase 2	PTGS2
36. Reverse primer			
37. Forward primer		Interleukin 8 receptor B	ILR8RB
38. Reverse primer			
39. Forward primer		Macrophage colony stimulating factor 1	CSF1 (MCSF1)
40. Reverse primer			
41. Forward primer		Melanoma growth stimulatory activity	MGSA
42. Reverse primer			
43. Forward primer		Chemokine (C-X-C motif) ligand 3	MGSA
44. Reverse primer			
45. Forward primer		Secreted phosphoprotein 1	SPP1 (OPN)
46. Reverse primer			
47. Forward primer		Cyclin D	CCND1
48. Reverse primer			
49. Forward primer		c-Myc	c-Myc
50. Reverse primer			

FIG. 1 (cont'd)

3/10

Sequence ID No. / ID	NCBI Entrez Database	Name	Abbreviation
51. Forward primer		Cyclin-dependent kinase inhibitor	HUMCDK1
52. Reverse primer			
53. Forward primer		Peroxisome proliferative activated receptor, alpha	PPAR $\alpha$
54. Reverse primer			
55. Forward primer		Peroxisome proliferative activated receptor, gamma	PPAR $\gamma$
56. Reverse primer			
57. Forward primer		Peroxisome proliferative activated receptor, delta	PPAR $\delta$
58. Reverse primer			
59. Forward primer		CD44 antigen	CD44
60. Reverse primer			
61. Forward primer		Prostaglandin-endoperoxide synthase 1	COX1
62. Reverse primer			
63. Forward primer		Serum amyloid A1	SAA1
64. Reverse primer			

FIG. 1 (cont'd)

4/10

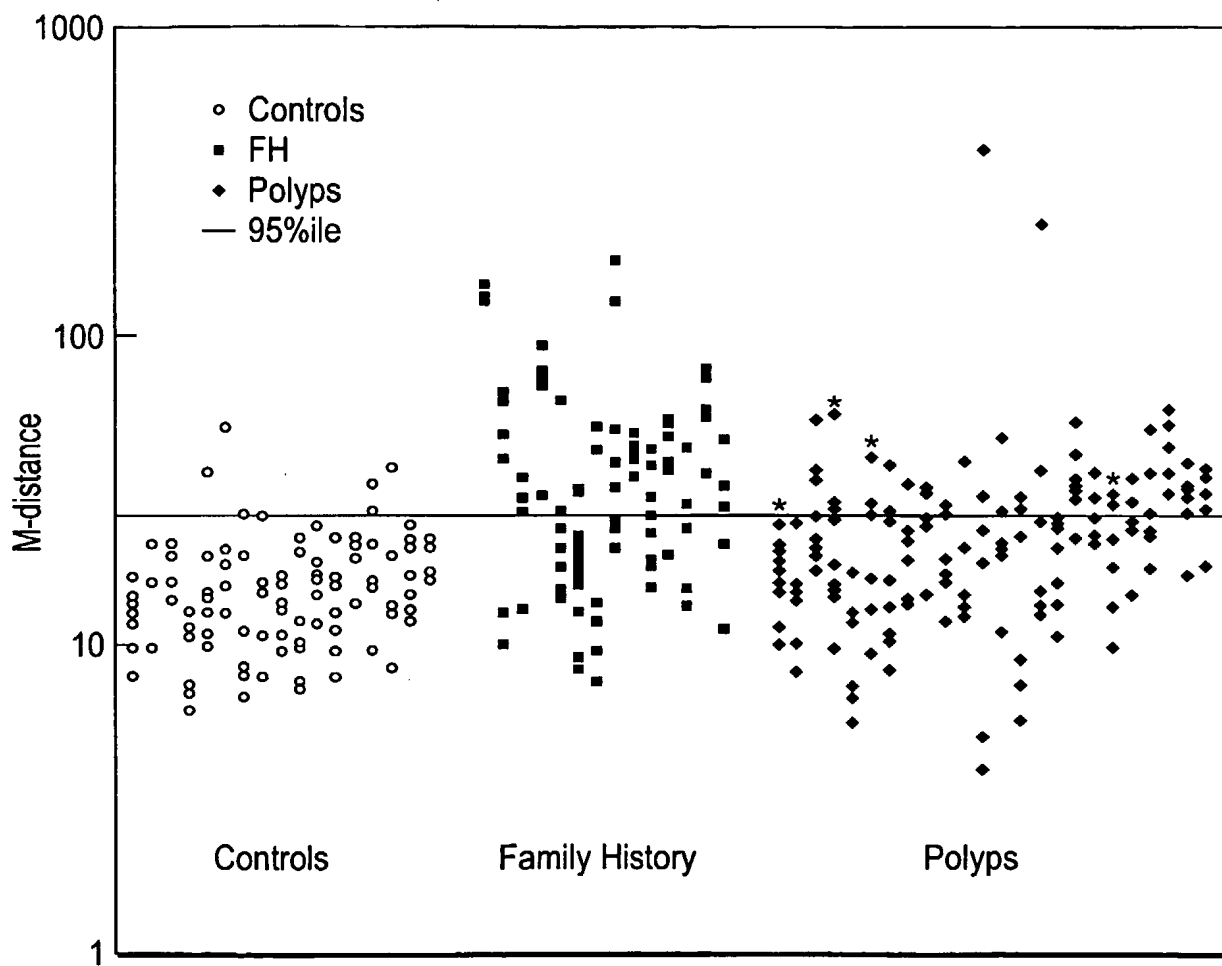
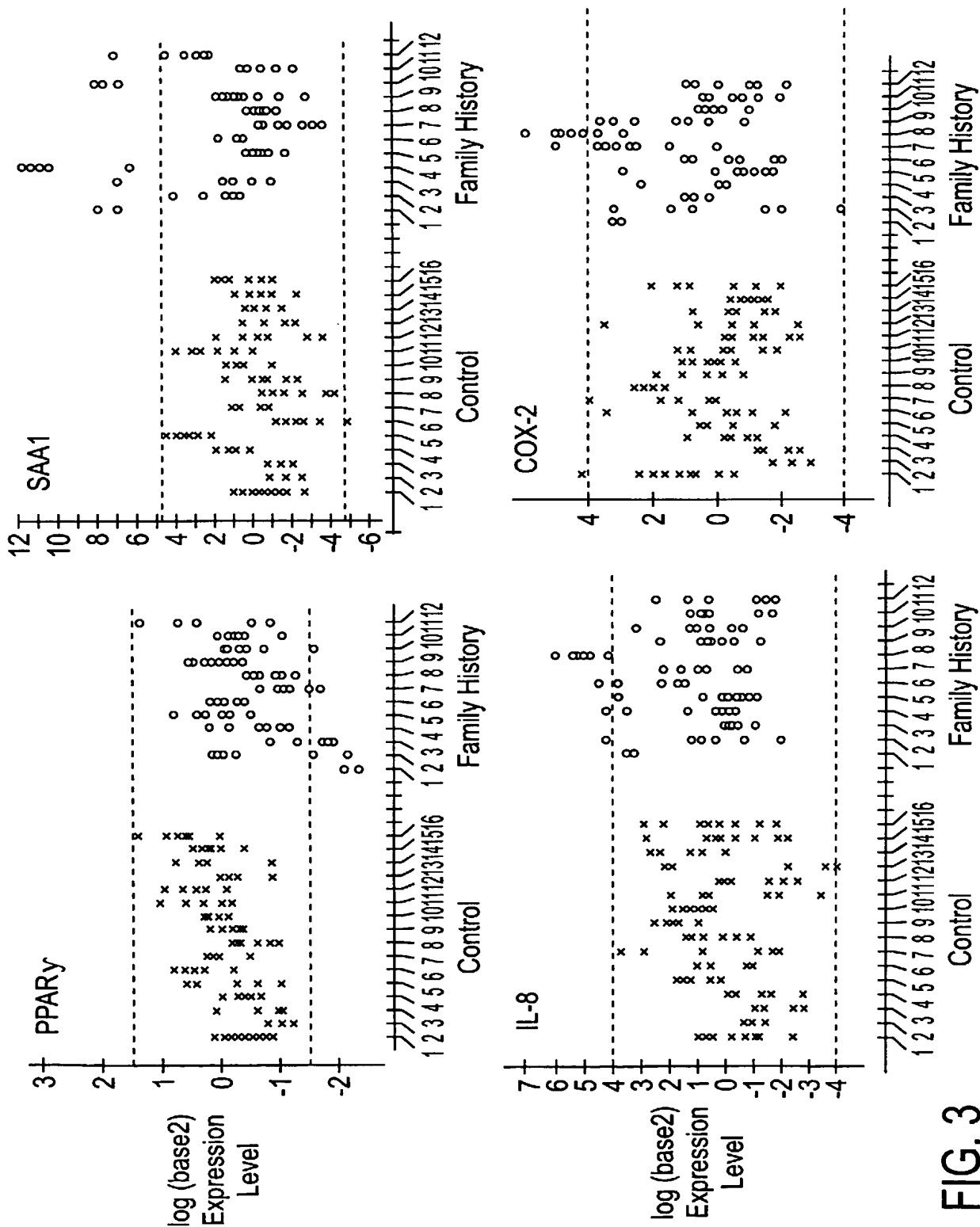


FIG. 2

5/10



### A) Rectosigmoid region

### Case 3 ( $n=5$ )

COX-2								
IL-8								
SAA1								
PPAR- $\gamma$								

Sigmoid → Rectum

### Case 4 ( $n=7$ )

COX-2									
IL-8									
SAA1									
PPAR- $\gamma$									

Sigmoid → Rectum

### Case 8 ( $n=8$ )

[illegible]

Sigmoid → Rectum

### B) Complete colon

### Case 3

	Asc	Trans	Des	S/R
Case 3				
COX-2				
IL-8				
SAA1				
PPAR- $\gamma$				

## Case 4

	Asc	Trans	Des	S/R
Case 4				
COX-2				
IL-8				
SAA1				
PPAR- $\gamma$				

## Case 8

	Asc	Trans	Des	S/R
Case 8				
COX-2				
IL-8				
SAA1				
PPAR- $\gamma$				

**FIG. 4**

7/10

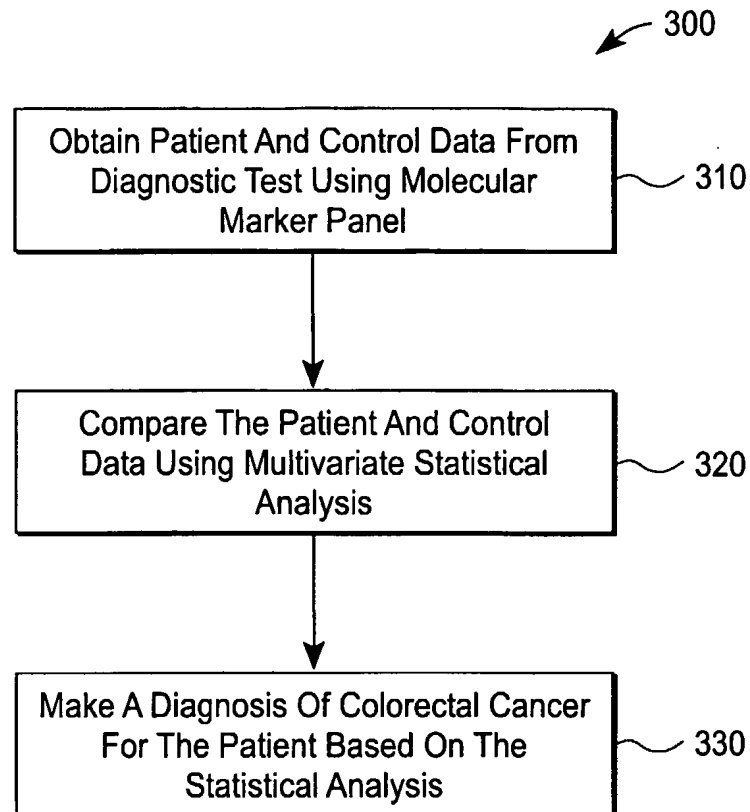


FIG. 5

8/10

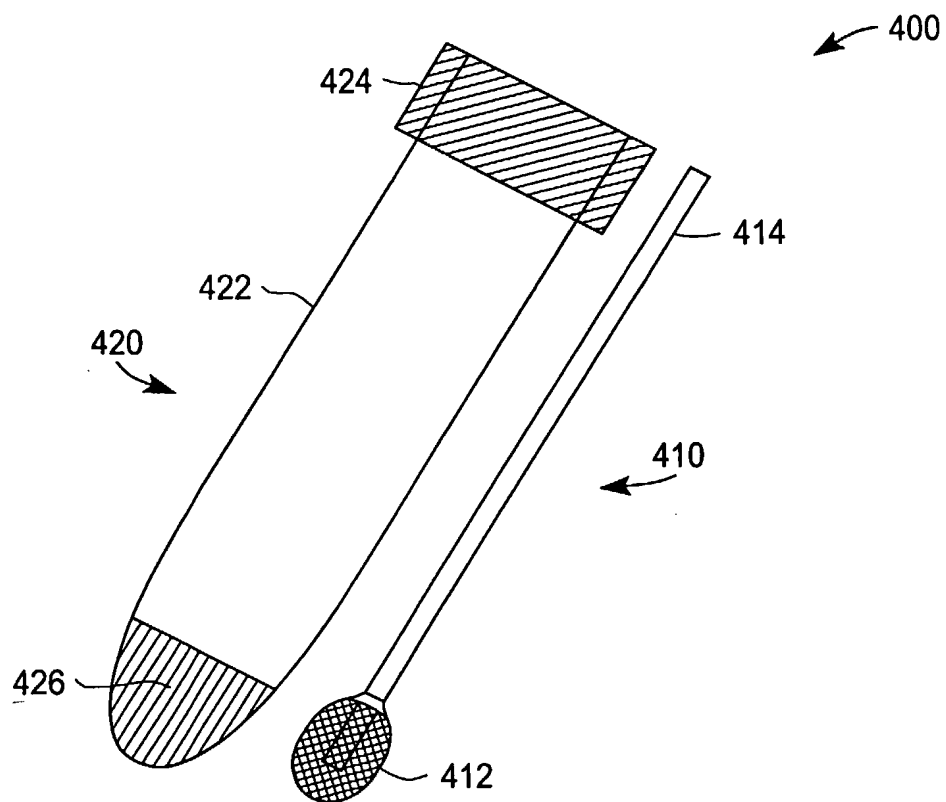


FIG. 6

9/10

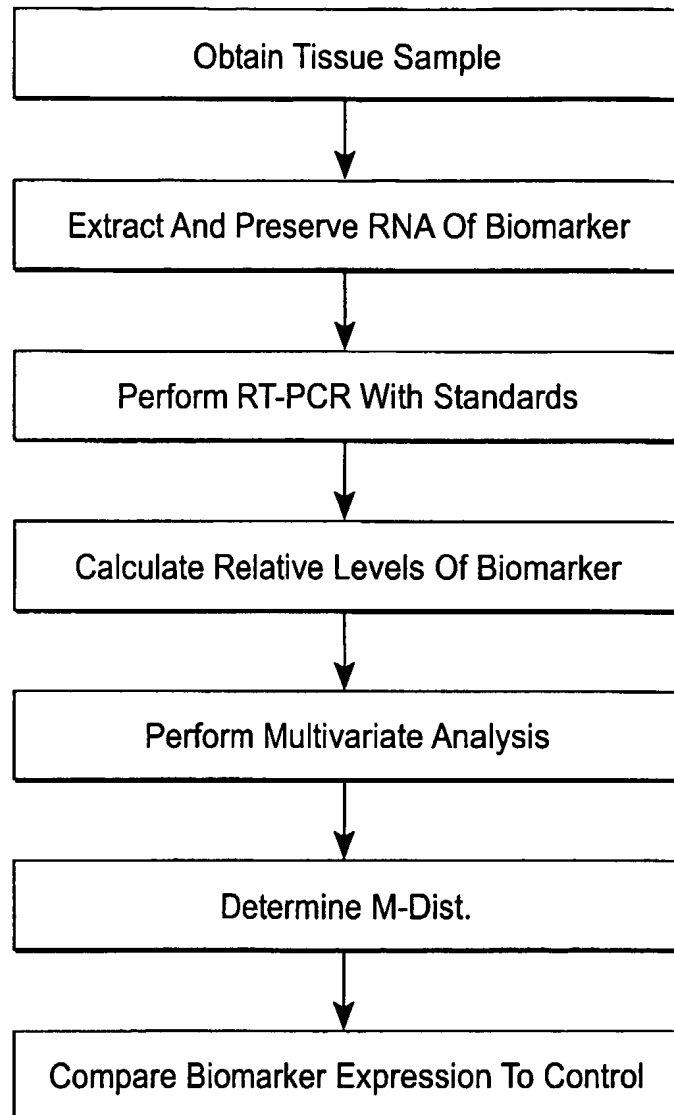


FIG. 7

10/10

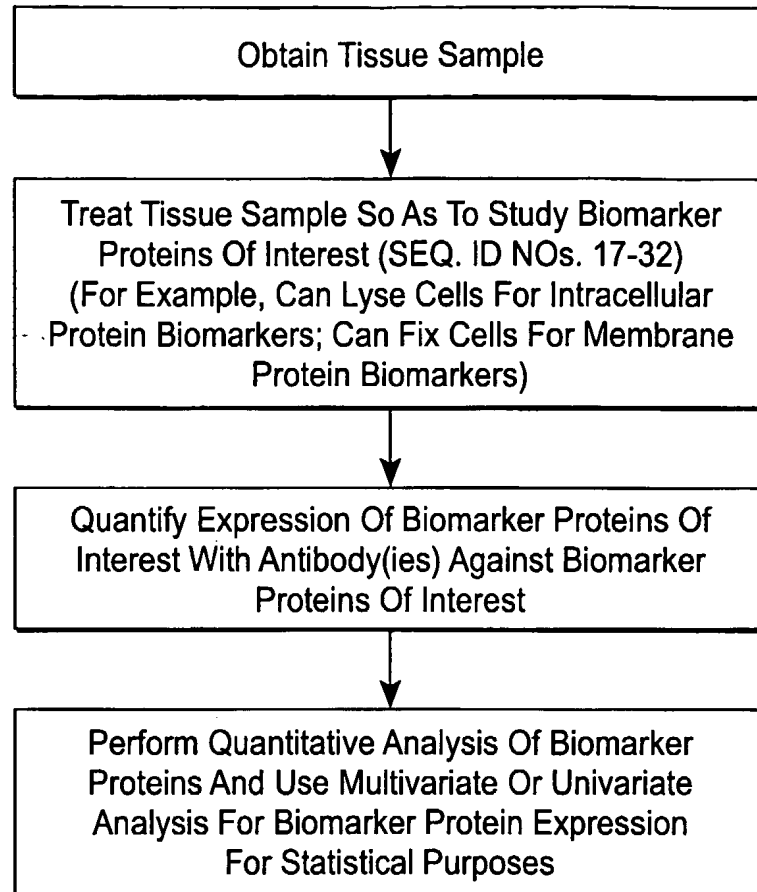


FIG. 8

## NLEE01001W00.ST25.txt

## EQUENCE LISTING

```

:110> Lee, Nancy M
:120> DRUG SCREENING AND MOLECULAR DIAGNOSTIC TEST FOR EARLY DETECTION
      OF COLORECTAL CANCER: REAGENTS, METHODS, AND KITS THEREOF
:130> NLEE-01001W00 MCF/MLB
:150> 60/614,746
:151> 2004-09-30
:150> 60/651,344
:151> 2005-02-08
:150> Not Assigned
:151> 2005-09-29
:160> 64
:170> PatentIn version 3.3
:210> 1
:211> 1629
:212> DNA
:213> HUMAN
:400> 1
jcagagcaca caagcttcta ggacaagagc caggaagaaa ccaccggaag gaaccatctc      60
ictgtgtgta aacatgactt ccaagctggc cgtggctctc ttggcagcct tcctgatttc      120
gcagctctg tgtgaagggtg cagttttgcc aaggagtgtc aaagaactta gatgtcagtg      180
ataaagaca tactccaaac ctttccaccc caaatttatc aaagaactga gagtgattga      240
jagtggacca cactgcgcca acacagaaat tatgtaaagc tttctgatgg aagagagctc      300
gtctggacc ccaaggaaaa ctgggtgcag agggttgtgg agaagttttt gaagagggct      360
jagaattcag aattcataaa aaaattcatt ctctgtggta tccaagaatc agtgaagatg      420
cagtgaaac ttcaagcaaa tctacttcaa cacttcatgt attgtgtggg tctgtttag      480
jgttgccaga tgcaatacaa gattcctggt taaatttgaa tttcagtaaa caatgaatag      540
ttttcattg taccatgaaa tatccagaac atacttatat gtaaagtatt atttatttga      600
ttctacaaaa aacaacaaat aattttttaa tataaggatt ttcctagata ttgcacggga      660
jaatatacaa atagcaaaat tgaggccaag ggccaagaga atatccgaac ttttaatttca      720
jgaattgaat gggtttgcta gaatgtgata tttgaagcat cacataaaaa tgatgggaca      780
ataaattttg ccataaagtc aaatttagct ggaaatcctg gatttttttc tgttaaactct      840
jgcaacccta gtctgctagc caggatccac aagtccttgt tccactgtgc cttggtttct      900
ctttatttcc taagtggaaa aagtattagc caccatctta cctcacagtg atgttgtgag      960

```

## NLEE01001W00.ST25.txt

gacatgtgga agcactttaa gttttttcat cataacataa attattttca agtgtaactt 1020  
 attaacctat ttattattta tgtattttatt taagcatcaa atattttgtgc aagaatttgg 1080  
 aaaaatagaa gatgaatcat tgattgaata gttataaaga tgttatagta aattttatttt 1140  
 atttttagata ttaaattgatg ttttattaga taaatttcaa tcagggtttt tagattaaac 1200  
 aaacaacaa ttgggtaccc agttaaattt tcatctcaga taaacaacaa ataatttttt 1260  
 agtataagta cattattggt tatctgaaat tttaattgaa ctaacaatcc tagtttgata 1320  
 ctcccagtct tgtcattgcc agctgtgttg gtagtgctgt gttgaattac ggaataatga 1380  
 gttagaacta ttaaaacagc caaaactcca cagtcaatat tagtaatttc ttgctggttg 1440  
 aaacttgttt attatgtaca aatagattct tataatatta tttaaattgac tgcattttta 1500  
 aatacaaggc tttatatttt taactttaag atgtttttat gtgctctcca aatttttttt 1560  
 actgtttctg attgtatgga aatataaaag taaatatgaa acatttaaaa tataatttgt 1620  
 tgtcaaagt 1629

<210> 2  
 <211> 3356  
 <212> DNA  
 <213> HUMAN

<400> 2  
 gtccaggaac tcctcagcag cgctccttc agctccacag ccagacgccc tcagacagca 60  
 aagcctaccc ccgcgccgcg ccctgcccgc cgctgcgatg ctgcgccgcg ccctgctgct 120  
 gtgcgcggtc ctggcgctca gccatacagc aaatccttgc tgttcccacc catgtcaaaa 180  
 ccgaggtgta tgtatgagtg tgggatttga ccagtataag tgcgattgta cccggacagg 240  
 attctatgga gaaaactgct caacaccgga atttttgaca agaataaaaat tatttctgaa 300  
 acccactcca aacacagtgc actacatact taccacttc aagggatttt ggaacgttgt 360  
 gaataacatt cccttccttc gaaatgcaat tatgagttat gtgttgacat ccagatcaca 420  
 tttgattgac agtccaccaa cttacaatgc tgactatggc taaaaagct gggaagcctt 480  
 ctctaacctc tcctattata ctagagccct tcctcctgtg cctgatgatt gcccgactcc 540  
 cttgggtgtc aaaggtaaaa agcagcttcc tgattcaaat gagattgtgg aaaaattgct 600  
 tctaagaaga aagttcatcc ctgatcccca gggctcaaac atgatgtttg cattctttgc 660  
 ccagcacttc acgcatcagt ttttcaagac agatcataag cgagggccag ctttcaccaa 720  
 cgggctgggc catgggggtgg acttaaataca tatttacggt gaaactctgg ctagacagcg 780  
 taaactgcgc cttttcaagg atggaaaaat gaaatatcag ataattgatg gagagatgta 840  
 tcctcccaca gtcaaagata ctcaggcaga gatgatctac cctcctcaag tccctgagca 900  
 tctacggttt gctgtggggc aggaggtctt tggctctgtg cctggtctga tgatgtatgc 960

NLEE01001w00.ST25.txt

cacaatctgg	ctgcgggaac	acaacagagt	atgcatgtg	cttaaacagg	agcatcctga	1020
atggggtgat	gagcagttgt	tccagacaag	caggctaata	ctgataggag	agactattaa	1080
gattgtgatt	gaagattatg	tgcaacactt	gagtggctat	cacttcaaac	tgaaatttga	1140
cccagaacta	cttttcaaca	aacaattcca	gtacccaaat	cgtattgctg	ctgaatttaa	1200
caccctctat	cactggcatc	cccttctgcc	tgacaccttt	caaattcatg	accagaaata	1260
caactatcaa	cagtttatct	acaacaactc	tatattgctg	gaacatggaa	ttaccagttt	1320
tgttgaatca	ttcaccaggc	aaattgctgg	caggggttgc	ggtggtagga	atgttcacc	1380
cgcagtacag	aaagtatcac	aggcttccat	tgaccagagc	aggcagatga	aataccagtc	1440
ttttaatgag	taccgcaaac	gctttatgct	gaagccctat	gaatcatttg	aagaacttac	1500
aggagaaaaag	gaaatgtctg	cagagttgga	agcactctat	ggtgacatcg	atgctgtgga	1560
gctgtatcct	gcccttctgg	tagaaaagcc	tcggccagat	gccatctttg	gtgaaaccat	1620
ggtagaagtt	ggagcaccat	tctccttgaa	aggacttatg	ggtaatgtta	tatgttctcc	1680
tgccactagg	aagccaagca	cttttggtgg	agaagtgggt	tttcaaatac	tcaacactgc	1740
ctcaattcag	tctctcatct	gcaataacgt	gaagggctgt	ccctttactt	cattcagtgt	1800
tccagatcca	gagctcatta	aaacagtcac	catcaatgca	agttcttccc	gctccggact	1860
agatgatatc	aatcccacag	tactactaaa	agaacgttcg	actgaactgt	agaagtctaa	1920
tgatcatatt	tattttattta	tatgaaccat	gtctattaat	ttaattattt	aataatattt	1980
atattaaact	ccttatgtta	cttaacatct	tctgtaacag	aagtcagtac	tcctgtttgcg	2040
gagaaaaggag	tcatacttgt	gaagactttt	atgtcactac	tctaaagatt	ttgctgttgc	2100
tgtaagttt	ggaaaacagt	ttttattctg	ttttataaac	cagagagaaa	tgagttttga	2160
cgtcttttta	cttgaatttc	aacttatatt	ataagaacga	aagtaaagat	gtttgaatac	2220
ttaaacactg	tcacaagatg	gcaaaatgct	gaaagttttt	acactgtcga	tgtttccaat	2280
gcatcttcca	tgatgcatta	gaagtaacta	atgtttgaaa	ttttaagta	cttttggtta	2340
tttttctgtc	atcaaacaaa	aacaggatc	agtgcattat	taaatgaata	tttaaatag	2400
acattaccag	taatttcatg	tctacttttt	aaaatcagca	atgaaacaat	aatttgaaat	2460
ttctaaattc	atagggtaga	atcacctgta	aaagcttggt	tgatttctta	aagttattaa	2520
actgtacat	ataccaaaaa	gaagctgtct	tggattttaa	tctgtaaaat	cagtagaaat	2580
tttactacaa	ttgcttggtta	aaatatttta	taagtgatgt	tcctttttca	ccaagagtat	2640
aaaccttttt	agtgtgactg	ttaaaacttc	cttttaaatc	aaaatgccaa	atttattaag	2700
gtgggtggagc	cactgcagtg	ttatcttaaa	ataagaatat	tttgttgaga	tattccagaa	2760
tttgtttata	tggtgtgtaa	catgtaaaat	ctatatcagc	aaaagggctc	acctttaaaa	2820
taagcaataa	caaagaagaa	aaccaaatta	ttgttcaaat	ttaggtttta	acttttgaag	2880

## NLEE01001WO0.ST25.txt

caaaccttttt	tttatccttg	tgcaactgcag	gcctggtact	cagatcttgc	tatgagggtta	2940
atgaagtacc	aagctgtgct	tgaataatga	tatgttttct	cagatcttct	gttgtacagt	3000
ttaatttagc	agtccatata	acattgcaaa	agtagcaatg	acctcataaa	atacctcttc	3060
aaaatgctta	aattcatttc	acacattaat	tttatctcag	tcttgaagcc	aattcagtag	3120
gtgcattgga	atcaagcctg	gctacctgca	tgctgttctt	tttcttttct	tcttttagcc	3180
atcttgctaa	gagacacagt	cttctcatca	cttcgtttct	cctatcttgt	tttactagtt	3240
ttaagatcag	agttcacttt	ctttggactc	tgcttatatt	ttcttacctg	aacttttgca	3300
agttttcagg	taaacctcag	ctcaggactg	ctatcttagct	cctcttaaga	agatta	3356

<210> 3  
 <211> 1750  
 <212> DNA  
 <213> HUMAN

<400> 3						
cctacagggtg	aaaagcccag	cgaccagtc	aggatttaag	tttacctcaa	aaatggaaga	60
ttttaacatg	gagagtga	gctttgaaga	tttctggaaa	ggtgaagatc	ttagtaatta	120
cagttacagc	tctaccctgc	cccccttct	actagatgcc	gccccatgtg	aaccagaatc	180
cctggaaatc	aacaagtatt	ttgtggtcat	tatctatgcc	ctgggtattcc	tgctgagcct	240
gctgggaaac	tcctctgtga	tgctggtcat	cttatacagc	agggtcggcc	gctccgtcac	300
tgatgtctac	ctgctgaacc	tagccttggc	cgacctactc	tttgccctga	ccttgcccat	360
ctgggcccgc	tccaagggtga	atggctggat	ttttggcaca	ttcctgtgca	agggtggtctc	420
actcctgaag	gaagtcaact	tctatagtgg	catcctgcta	ctggcctgca	tcagtgtgga	480
ccgttacctg	gccattgtcc	atgccacacg	cacactgacc	cagaagcgct	acttgggtcaa	540
attcatatgt	ctcagcatct	gggggtctgtc	cttgctcctg	gccctgcctg	tcttactttt	600
ccgaaggacc	gtctactcat	ccaatgttag	cccagcctgc	tatgaggaca	tgggcaacaa	660
tacagcaaac	tggcggatgc	tgttacggat	cctgccccag	tcctttggct	tcctcgtgcc	720
actgctgatc	atgctgttct	gctacggatt	caccctgcgt	acgctgttta	aggcccacat	780
ggggcagaag	caccgggcca	tgcggtcat	ctttgctgtc	gtcctcatct	tcctgctttg	840
ctggctgccc	tacaacctgg	tcctgctggc	agacaccctc	atgaggaccc	agggtgatcca	900
ggagacctgt	gagcgccgca	atcacatcga	ccgggctctg	gatgccaccg	agattctggg	960
catccttcac	agctgcctca	acccctcat	ctacgccttc	attggccaga	agtttcgcca	1020
tggactcctc	aagattctag	ctatacatgg	cttgatcagc	aaggactccc	tgcccaaaga	1080
cagcaggcct	tcctttgttg	gctcttcttc	agggcacact	tccactactc	tctaagacct	1140
cctgcctaag	tgacgccccg	tgggggttct	cccttctctt	cacagtcaca	ttccaagcct	1200

## NLEE01001WO0.ST25.txt

catgtccact	ggttcttctt	ggtctcagtg	tcaatgcagc	ccccattgtg	gtcacaggaa	1260
gcagaggagg	ccacgttctt	actagtttcc	cttgcattgt	ttagaaagct	tgccctggtg	1320
cctcaccctt	tgccataatt	actatgtcat	ttgctggagc	tctgcccata	ctgcccctga	1380
gcccattggc	ctctatgttc	taagaagtga	aaatctacac	tccagtgaga	cagctctgca	1440
tactcattag	gatggctagt	atcaaaagaa	agaaaatcag	gctggccaac	gggatgaaac	1500
cctgtctcta	ctaaaaatac	aaaaaaaaaa	aaaaaatta	gccgggctg	gtggtgagtg	1560
cctgtaatac	cagctacttg	ggaggctgag	atgggagaa	cacttgaacc	cgggaggcag	1620
aggttgagc	gagccgagat	tgtgcccctg	cactccagcc	tgagcgacag	tgagactctg	1680
tctcagtgca	tgaagatgta	gaggagaaac	tggaactctc	gagcgttgct	gggggggatt	1740
gtaaaatggt						1750

<210> 4  
 <211> 3939  
 <212> DNA  
 <213> HUMAN

<400> 4						
cctgggtcct	ctcggcgcca	gagccgctct	ccgcatccca	ggacagcggg	gcggccctcg	60
gccggggcgc	ccactccgca	gcagccagcg	agccagctgc	cccgtatgac	cgcgccgggc	120
gccgcccggc	gctgccctcc	cacgacatgg	ctgggctccc	tgctgttggt	ggtctgtctc	180
ctggcgagca	ggagtatcac	cgaggagggtg	tcggagtact	gtagccacat	gattgggagt	240
ggacacctgc	agtctctgca	gcggctgatt	gacagtcaga	tggagacctc	gtgccaaatt	300
acatttgagt	ttgtagacca	ggaacagttg	aaagatccag	tgtgctacct	taagaaggca	360
tttctcctgg	tacaagacat	aatggaggac	accatgcgct	tcagagataa	caccgccaat	420
cccatcgcca	ttgtgcagct	gcaggaactc	tctttgaggc	tgaagagctg	cttcaccaag	480
gattatgaag	agcatgacaa	ggcctgcgtc	cgaactttct	atgagacacc	tctccagttg	540
ctggagaagg	tcaagaatgt	ctttaatgaa	acaaagaatc	tccttgacaa	ggactggaat	600
attttcagca	agaactgcaa	caacagcttt	gctgaatgct	ccagccaaga	tgtggtgacc	660
aagcctgatt	gcaactgcct	gtaccccaaa	gccatcccta	gcagtgaccc	ggcctctgtc	720
tcccctcatc	agcccctcgc	cccctccatg	gcccctgtgg	ctggcttgac	ctgggaggac	780
tctgagggaa	ctgagggcag	ctccctcttg	cctggtgagc	agcccctgca	cacagtggat	840
ccaggcagtg	ccaagcagcg	gccaccagcg	agcacctgcc	agagctttga	gccgccagag	900
accccagttg	tcaaggacag	caccatcggt	ggctcaccac	agcctcgccc	ctctgtcggg	960
gccttcaacc	ccgggatgga	ggatattctt	gactctgcaa	tgggcactaa	ttgggtccca	1020
gaagaagcct	ctggagaggc	cagtgaagatt	cccgtacccc	aaggacaga	gctttcccc	1080

## NLEE01001WO0.ST25.txt

tccaggccag	gagggggcag	catgcagaca	gagcccgcc	gacccagcaa	cttcctctca	1140
gcattcttctc	cactccctgc	atcagcaaag	ggccaacagc	cggcagatgt	aactgctaca	1200
gccttgccca	gggtgggccc	cgtgatgcc	actggccagg	actggaatca	cacccccag	1260
aagacagacc	atccatctgc	cctgctcaga	gacccccgg	agccaggctc	tcccaggatc	1320
tcatcactgc	gccccaggc	cctcagcaac	ccctccaccc	tctctgctca	gccacagctt	1380
tccagaagcc	actcctcggg	cagcgtgctg	ccccttgggg	agctggaggg	caggaggagc	1440
accagggatc	ggacgagccc	cgcagagcca	gaagcagcac	cagcaagtga	aggggcagcc	1500
aggcccctgc	cccgttttaa	ctccgttcct	ttgactgaca	caggccatga	gaggcagtcc	1560
gagggatcct	ccagcccgc	gctccaggag	tctgtcttcc	acctgctggt	gcccagtgtc	1620
atcctggtct	tgctggctgt	cggaggcctc	ttgttctaca	ggtggaggcg	gcggagccat	1680
caagagcctc	agagagcgga	ttctcccttg	gagcaaccag	agggcagccc	cctgactcag	1740
gatgacagac	aggtggaact	gccagtgtag	agggaattct	aagctggacg	cacagaacag	1800
tctcttcgtg	ggaggagaca	ttatggggcg	tccaccacca	cccctccctg	gccatcctcc	1860
tggaatgtgg	tctgccctcc	accagagctc	ctgcctgcc	ggactggacc	agagcagcca	1920
ggctggggcc	cctctgtctc	aaccgcaga	cccttgactg	aatgagagag	gccagaggat	1980
gctccccatg	ctgccactat	ttattgtgag	ccctggaggc	tcccatgtgc	ttgaggaagg	2040
ctggtgagcc	cggctcagga	ccctcttccc	tcaggggctg	cagcctcctc	tactccctt	2100
ccatgccgga	accaggcca	gggacccacc	ggcctgtggt	ttgtgggaaa	gcagggtgca	2160
cgctgaggag	tgaacaacc	ctgcacccag	agggcctgcc	tggtgccaag	gtatcccagc	2220
ctggacaggc	atggacctgt	ctccagacag	aggagcctga	agttcgtggg	gcgggacagc	2280
ctcggcctga	tttcccgtaa	aggtgtgcag	cctgagagac	gggaagagga	ggcctctgca	2340
cctgctggtc	tgcactgaca	gcctgaaggg	tctacaccct	cggctcacct	aagtcctgt	2400
gctggttgcc	aggcccagag	gggaggccag	ccctgccctc	aggacctgcc	tgacctgcca	2460
gtgatgccaa	gagggggatc	aagcactggc	ctctgcccct	cctccttcca	gcacctgcca	2520
gagcttctcc	agcaggccaa	gcagaggctc	ccctcatgaa	ggaagccatt	gcactgtgaa	2580
cactgtacct	gcctgctgaa	cagcctcccc	ccgtccatcc	atgagccagc	atccgtccgt	2640
cctccactct	ccagcctctc	cccagcctcc	tgcactgagc	tggcctcacc	agtcgactga	2700
gggagcccct	cagccctgac	cttctcctga	cctggccttt	gactccccgg	agtggagtgg	2760
ggtgggagaa	cctcctgggc	cgccagccag	agccgtcttt	taggctgtgt	tcttcgcccc	2820
ggtttctgca	tcttccactt	tgacattccc	aagaggggaag	ggactagtgg	gagagagcaa	2880
gggaggggag	ggcacagaca	gagagcctac	agggcgagct	ctgactgaag	atgggccttt	2940

## NLEE01001W00.ST25.txt

gaaatatagg tatgcacctg aggttggggg aggggtctgca ctcccaaacc ccagcgcagt 3000  
 gtcctttccc tgctgccgac aggaacctgg ggctgagcag gttatccctg tcaggagccc 3060  
 tggactgggc tgcattctcag cccacactgc atgggtatcca gctcccatcc acttctcacc 3120  
 cttctttcct cctgaccttg gtcagcagtg atgacctcca actctcacc accccctcta 3180  
 ccatcacctc taaccaggca agccagggtg ggagagcaat caggagagcc aggcctcagc 3240  
 ttccaatgcc tggagggcct ccactttgtg gccagcctgt ggtgctggct ctgaggccta 3300  
 ggcaacgagc gacagggctg ccagttgccc ctgggttcct ttgtgctgct gtgtgcctcc 3360  
 tctcctgccg ccctttgtcc tccgctaaga gaccctgccc tacctggccg ctgggccccg 3420  
 tgactttccc ttcttgccca ggaaagttag ggtcggtgg cccaccttc cctgtcctga 3480  
 tgccgacagc ttagggaagg gcactgaact tgcatatggg gcttagcctt ctagtcacag 3540  
 cctctatatt tgatgctaga aaacacatat ttttaaatgg aagaaaaata aaaaggcatt 3600  
 ccccttcat cccctacct taaacatata atattttaaa ggtcaaaaaa gcaatccaac 3660  
 cactgcaga agctcttttt gagcacttgg tggcatcaga gcaggaggag cccagagcc 3720  
 acctctggtg tccccaggc tacctgctca ggaacccctt ctgttctctg agaactcaac 3780  
 agaggacatt ggctcacgca ctgtgagatt ttgtttttat acttgcaact ggtgaattat 3840  
 tttttataaa gtcatttaaa tatctattta aaagatagga agctgcttat atatttaata 3900  
 ataaaagaag tgcacaagct gccgttgacg tagctcgag 3939

<210> 5  
 <211> 1024  
 <212> DNA  
 <213> HUMAN

<400> 5  
 atggcccgcg ctgctctctc cgccgcccc agcaatcccc ggctcctgcg agtggcactg 60  
 ctgctcctgc tcctggtagc cgctggccgg cgcgagcag gagcgtccgt ggccactgaa 120  
 ctgctgctgc agtgcttgca gaccctgcag ggaattcacc ccaagaacat ccaaagtgtg 180  
 aacgtgaagt ccccgacc ccactgcgcc caaacgaag tcatagccac actcaagaat 240  
 gggcggaag cttgcctcaa tcctgcatcc cccatagtta agaaaatcat cgaaaagatg 300  
 ctgaacagtg acaaattcaa ctgaccagaa gggaggagga agctcactgg tggctgttcc 360  
 tgaaggaggc cctgccctta taggaacaga agaggaaaga gagacacagc tgcagaggcc 420  
 acctggattg tgcctaattg gtttgagcat cgcttaggag aagtcttcta tttatttatt 480  
 tattcattag ttttgaagat tctatgttaa tatttttaggt gtaaaataat taagggtatg 540  
 attaactcta cctgcacact gtcctattat attcattctt tttgaaatgt caacccaag 600  
 ttagttcaat ctggattcat atttaatttg aaggtagaat gttttcaaat gttctccagt 660

NLEE01001WO0.ST25.txt

cattatgtta atatttctga ggagcctgca acatgccagc cactgtgata gaggctggcg	720
gatccaagca aatggccaat gagatcattg tgaaggcagg ggaatgtatg tgcacatctg	780
ttttgtaact gtttagatga atgtcagttg ttattttattg aaatgatattc acagtgtgtg	840
gtcaacattt ctcatgttga aactttaaga actaaaatgt tctaaatatc ccttggacat	900
tttatgtctt tcttgtaagg catactgcct tgtttaatgg tagttttaca gtgtttctg	960
cttagaaca aggggcttaa ttattgatgt ttcatagag aatataaaaa taaagcactt	1020
atag	1024

<210> 6  
 <211> 1064  
 <212> DNA  
 <213> HUMAN

<220>  
 <221> misc\_feature  
 <222> (27)..(27)  
 <223> n = a, c, g, t

<220>  
 <221> misc\_feature  
 <222> (766)..(766)  
 <223> n = a, c, g, t

<400> 6	
cacagccggg tcgcaggcac ctccccngcc agctctcccg cattctgcac agcttcccga	60
cgcgctctgct gagcccatg gccacgcca cgctctccgc cgccccagc aatccccggc	120
tcctgcgggt ggcgctgctg ctctgctcc tgggtgggcag ccggcgcgca gcaggagcgt	180
ccgtgggtcac tgaactgcgc tgccagtgtg tgcagacact gcagggaatt cacctcaaga	240
acatccaaag tgtgaatgta aggtcccccg gaccccaactg cgcccaaacc gaagtcatag	300
ccacactcaa gaatgggaag aaagcttgct tcaaccccg atcccccatg gttcagaaaa	360
tcactgaaaa gatactgaac aaggggagca ccaactgaca ggagagaagt aagaagctta	420
tcagcgtatc attgacactt cctgcagggt ggtccctgcc ctaccagag ctgaaaatga	480
aaaagagaac agcagctttc tagggacagc tggaaaggga cttaatgtgt ttgactat	540
cttacgaggg ttctacttat ttatgtat	600
catgctgtta tttaaagatg tgagtgtgtt tcatcaaaca tagctcagtc ctgattat	660
aattggaata tgatggggtt taaatgtgtc attaaactaa tatttagtgg gagaccataa	720
tgtgtcagcc accttgataa atgacagggt ggggaactgg agggtngggg gattgaaatg	780
caagcaatta gtggatcact gttagggttaa gggaatgtat gtacacatct atttttata	840
cttttttttt taaaaaagaa tgtcagttgt tatttattca aattatctca cattatgtgt	900
tcaacatttt tatgctgaag tttcccttag acattttatg tcttgcttgt agggcataat	960

## NLEE01001wo0.ST25.txt

gccttgttta atgtccattc tgcagcgttt ctctttccct tggaaaagag aatttatcat 1020  
tactgtttaca tttgtacaaa tgacatgata ataaaagtgt tatg 1064

<210> 7  
<211> 1469  
<212> DNA  
<213> HUMAN

<400> 7  
agcagcagga ggaggcagag cacagcatcg tcgggaccag actcgtctca ggccagttgc 60  
agcctttctca gccaaacgcc gaccaaggaa aactcactac catgagaatt gcagtgattt 120  
gcttttgcct cctaggcatc acctgtgcc aaccaggttaa acaggctgat tctggaagtt 180  
ctgaggaaaa gcagctttac aacaaatacc cagatgctgt ggccacatgg ctaaaccctg 240  
acccatctca gaagcagaat ctcttagccc cacagaccct tccaagtaag tccaacgaaa 300  
gccatgacca catggatgat atggatgatg aagatgatga tgaccatgtg gacagccagg 360  
actccattga ctcgaacgac tctgatgatg tagatgacac tgatgattct caccagtctg 420  
atgagtctca ccattctgat gaatctgatg aactgggtcac tgattttccc acggacctgc 480  
cagcaaccga agttttcact ccagttgtcc ccacagtaga cacatatgat ggccgaggtg 540  
atagtgtggt ttatggactg aggtcaaaat ctaagaagtt tcgcagacct gacatccagt 600  
accctgatgc tacagacgag gacatcacct cacacatgga aagcgaggag ttgaatggtg 660  
catacaaggc catccccgtt gcccaggacc tgaacgcgcc ttctgattgg gacagccgtg 720  
ggaaggacag ttatgaaacg agtcagctgg atgaccagag tgctgaaacc cacagccaca 780  
agcagtccag attatataag cggaaagcca atgatgagag caatgagcat tccgatgtga 840  
ttgatagtca ggaactttcc aaagtcagcc gtgaattcca cagccatgaa tttcacagcc 900  
atgaagatat gctggttgta gaccccaaaa gtaaggaaga agataaacac ctgaaatttc 960  
gtattttctca tgaattagat agtgcattct ctgagggtcaa ttaaaaggag aaaaaataca 1020  
atttctcact ttgcatttag tcaaaagaaa aaatgcttta tagcaaatg aaagagaaca 1080  
tgaaatgctt ctttctcagt ttattggttg aatgtgtatc tatttgagtc tggaaataac 1140  
taatgtgttt gataattagt ttagtttggt gcttcatgga aactccctgt aaactaaaag 1200  
cttcagggtt atgtctatgt tcattctata gaagaaatgc aaactatcac tgtattttta 1260  
tatttgttat tctctcatga atagaaattt atgtagaagc aaacaaaata cttttaccca 1320  
cttaaaaaga gaatataaca ttttatgtca ctataatctt ttgtttttta agttagtgtg 1380  
tattttgttg tgattatctt tttgtggtgt gaataaatct tttatcttga atgtaataag 1440  
aaaaaaaaa aaaaaaCa aa aaaaaaa 1469

## NLEE01001W00.ST25.txt

<210> 8  
 <211> 1256  
 <212> DNA  
 <213> HUMAN

<400> 8  
 gcagtagcag cgagcagcag agtccgcacg ctccggcgag gggcagaaga gcgcgagggg 60  
 gcgcgggggca gcagaagcga gagccgagcg cggacccagc caggaccac agccctcccc 120  
 agctgcccag gaagagcccc agccatggaa caccagctcc tgtgctgcga agtggaacc 180  
 atccgcccgcg cgtaccccgga tgccaacctc ctcaacgacc ggggtgctgcg ggccatgctg 240  
 aaggcggagg agacctgcgc gccctcggtg tcctacttca aatgtgtgca gaaggaggtc 300  
 ctgccgtcca tgcggaagat cgtcgccacc tggatgctgg aggtctgcga ggaacagaag 360  
 tgcgaggagg aggtcttccc gctggccatg aactacctgg accgcttcct gtcgctggag 420  
 cccgtgaaaa agagccgcct gcagctgctg ggggccactt gcatgttcgt ggcctctaag 480  
 atgaaggaga ccatccccct gacggccgag aagctgtgca tctacaccga cggctccatc 540  
 cggcccagagg agctgctgca aatggagctg ctcttggtga acaagctcaa gtggaacctg 600  
 gccgcaatga ccccgcacga tttcattgaa cacttcctct ccaaaatgcc agaggcggag 660  
 gagaacaaac agatcatccg caaacacgcg cagaccttcg ttgcctcttg tgccacagat 720  
 gtgaagtcca tttccaatcc gccctccatg gtggcagcgg ggagcgtggt ggccgcagtg 780  
 caaggcctga acctgaggag ccccaacaac ttcctgtcct actaccgcct cacacgcttc 840  
 ctctccagag tgatcaagtg tgaccagac tgcctccggg cctgccagga gcagatcgaa 900  
 gccctgctgg agtcaagcct gcgccaggcc cagcagaaca tggaccccaa ggccgccgag 960  
 gaggaggaag aggaggagga ggagggtggac ctggcttgca caccaccga cgtgcgggac 1020  
 gtggacatct gagggggcca ggcaggcggg cgccaccgcc accgcagcg agggcggagc 1080  
 cggccccagg tgctccacat gacagtccct cctctccgga gcattttgat accagaaggg 1140  
 aaagcttcat tctccttggt gttggttggt ttttccttg ctctttcccc cttccatctc 1200  
 tgacttaagc aaaagaaaaa gattacccaa aaactgtctt taaaagagag agagag 1256

<210> 9  
 <211> 2121  
 <212> DNA  
 <213> HUMAN

<400> 9  
 ctgctcgcgg ccgccaccgc cgggccccgg ccgtccctgg ctcccctcct gcctcgagaa 60  
 gggcagggct tctcagaggc ttggcgggaa aaaagaacgg agggagggat cgcgctgagt 120  
 ataaaagccg gttttcgggg ctttatctaa ctcgctgtag taattccagc gagaggcaga 180  
 gggagcgagc gggcggccgg ctaggggtga agagccgggc gagcagagct gcgctgcggg 240

NLEE01001W00.ST25.txt

cgctcctggga aggggagatcc ggagcgaata gggggccttcg cctctggccc agccctcccc	300
cttgatcccc caggccagcg gtccgcaacc cttgccgcat ccacgaaact ttgcccatag	360
cagcgggcg gcaactttgca ctggaactta caacacccga gcaaggacgc gactctcccc	420
acgcggggag gctattctgc ccatttgggg acacttcccc gccgctgcca ggacccgctt	480
ctctgaaagg ctctccttgc agctgcttag acgctggatt tttttcgggt agtggaaaac	540
cagcagcctc ccgcgacgat gcccctcaac gttagcttca ccaacaggaa ctatgacctc	600
gactacgact cggtgcagcc gtattttctac tgcgacgagg aggagaactt ctaccagcag	660
cagcagcaga gcgagctgca gccccggcg cccagcgagg atatctggaa gaaattcgag	720
ctgctgcccc ccccgccctt gtcccctagc cgccgctccg ggctctgctc gccctcctac	780
gttgcggtca cacccttctc ctttcgggga gacaacgacg gcggtggcgg gagcttctcc	840
acggccgacc agctggagat ggtgaccgag ctgctgggag gagacatggt gaaccagagt	900
ttcatctgcy acccgagcga cgagaccttc atcaaaaaca tcatcatcca ggactgtatg	960
tggagcggct tctcgccgc cgccaagctc gtctcagaga agctggcctc ctaccaggct	1020
gcgcgcaaag acagcggcag cccgaacccc gcccgcggcc acagcgtctg ctccacctcc	1080
agcttgtacc tgcaggatct gagcgccgcc gcctcagagt gcatcgacct ctcggtggtc	1140
ttcccctacc ctctcaacga cagcagctcg cccaagtctt gcgcctcgca agactccagc	1200
gccttctctc cgtcctcgga ttctctgctc tcctcgacgg agtcctcccc gcagggcagc	1260
cccgagcccc tgggtgtcca tgaggagaca ccgccacca ccagcagcga ctctgaggag	1320
gaacaagaag atgaggaaga aatcgatgtt gtttctgtgg aaaagaggca ggctcctggc	1380
aaaagggtcag agtctggatc accttctgct ggaggccaca gcaaacctcc tcacagccca	1440
ctggtcctca agaggtgcca cgtctccaca catcagcaca actacgcagc gcctccctcc	1500
actcggaagg actatcctgc tgccaagagg gtcaagttag acagtgtcag agtcctgaga	1560
cagatcagca acaaccgaaa atgcaccagc cccaggtcct cggacaccga ggagaatgtc	1620
aagaggcgaa cacacaacgt cttggagcgc cagaggagga acgagctaaa acggagcttt	1680
tttgccctgc gtgaccagat cccggagttg gaaaacaatg aaaaggcccc caaggtagtt	1740
atccttaaaa aagccacagc atacatcctg tccgtccaag cagaggagca aaagctcatt	1800
tctgaagagg acttgttgcy gaaacgacga gaacagttga aacacaaact tgaacagcta	1860
cggaactctt gtgcgtaagg aaaagtaagg aaaacgattc cttctaacag aaatgtcctg	1920
agcaatcacc tatgaacttg tttcaaatgc atgatcaaat gcaacctcac aaccttggtc	1980
gagtcttgag actgaaagat ttagccataa tgtaaaactgc ctcaaattgg actttgggca	2040
taaaagaact tttttatgct taccatcttt ttttttctt taacagattt gtatttaaga	2100
attgttttta aaaaatttta a	2121

## NLEE01001WO0.ST25.txt

<210> 10  
 <211> 2098  
 <212> DNA  
 <213> HUMAN

<400> 10  
 cctgccgaag tcagttcctt gtggagccgg agctgggagc ggattcgccg aggcaccgag 60  
 gcactcagag gaggcgccat gtcagaaccg gctggggatg tccgtcagaa cccatgcggc 120  
 agcaaggcct gccgccgctt cttcggccca gtggacagcg agcagctgag ccgcgactgt 180  
 gatgcgctaa tggcgggctg catccaggag gcccgtagag gatggaactt cgactttgtc 240  
 accgagacac cactggaggg tgacttcgcc tgggagcgtg tgcggggcct tggcctgccc 300  
 aagctctacc ttccacggg gccccggcga ggccgggatg agttgggagg aggcaggcgg 360  
 cctggcacct cacctgctct gctgcagggg acagcagagg aagaccatgt ggacctgtca 420  
 ctgtcttgta cccttgctgc tcgctcaggg gagcaggctg aagggtcccc aggtggacct 480  
 ggagactctc agggtcgaaa acggcggcag accagcatga cagatttcta ccactccaaa 540  
 cgccggctga tcttctccaa gaggaagccc taatccgccc acaggaagcc tgcagtcctg 600  
 gaagcgcgag ggcctcaaag gcccgctcta catcttctgc cttagtctca gtttggtgtg 660  
 cttaattatt atttgtgttt taatttaaac acctcctcat gtacataccc tggccgcccc 720  
 ctgcccccca gcctctggca ttagaattat ttaaacaaaa actaggcggg tgaatgagag 780  
 gttcctaaga gtgctgggca tttttatatt atgaaatact atttaaagcc tcctcatccc 840  
 gtgttctcct tttcctctct cccggagggt gggtagggcc gcttcatgcc agctacttcc 900  
 tcctccccac ttgtccgctg ggtggtaccc tctggagggg tgtggctcct tcccatcgct 960  
 gtcacaggcg gttatgaaat tcacccccct tcctggacac tcagacctga attctttttc 1020  
 atttgagaag taaacagatg gcactttgaa ggggcctcac cgagtggggg catcatcaaa 1080  
 aactttggag tcccctcacc tcctctaagg ttgggcaggg tgaccctgaa gtgagcacag 1140  
 cctagggctg agctggggac ctggtaccct cctggctctt gataaccccc tctgtcttgt 1200  
 gaaggcaggg ggaagggtgg gtcctggagc agaccacccc gcctgccctc atggccccctc 1260  
 tgacctgcac tggggagccc gtctcagtgt tgagcctttt ccctctttgg ctccccgtga 1320  
 ccttttgagg agccccagct acccttcttc tcagctggg ctctgcaatt cccctctgct 1380  
 gctgtccctc ccccttgctc tttcccttca gtaccctctc agctccagggt ggctctgagg 1440  
 tgcctgtccc accccacccc ccagctcaat ggactggaag gggaaggag acacaagaag 1500  
 aagggcaccc tagttctacc tcaggcagct caagcagcga ccgccccctc ctctagctgt 1560  
 gggggtgagg gtcccatgtg gtggcacagg ccccttgag tggggttatc tctgtgttag 1620  
 gggatatatga tgggggagta gatctttcta ggaggagac actggcccct caaatcgtcc 1680

## NLEE01001W00.ST25.txt

agcgaccttc	ctcatccacc	ccatccctcc	ccagttcatt	gcactttgat	tagcagcgga	1740
acaaggagtc	agacatttta	agatggtggc	agtagaggct	atggacaggg	catgccacgt	1800
gggctcatat	ggggctggga	gtagtgtct	ttcctggcac	taacgttgag	cccctggagg	1860
cactgaagtg	cttagtgtac	ttggagtatt	ggggtctgac	cccaaacacc	ttccagctcc	1920
tgtaacatac	tggcctggac	tgttttctct	cggctcccca	tgtgtcctgg	ttcccgtttc	1980
tccacctaga	ctgtaaacct	ctcgagggca	gggaccacac	cctgtactgt	tctgtgtctt	2040
tcacagctcc	tcccacaatg	ctgatataca	gcaggtgctc	aataaacgat	tcttagtg	2098

<210> 11  
 <211> 1850  
 <212> DNA  
 <213> HUMAN

<400> 11	
ggcccaggct	gaagctcagg gccctgtctg ctctgtggac tcaacagttt gtggcaagac 60
aagctcagaa	ctgagaagct gtcaccacag ttctggaggc tgggaagtgc aagatcaaag 120
tgccagcaga	ttcagtgtca tgtgaggacg tgcttcctgc ttcataagata agagcttgga 180
gctcggcgca	caaccagcac catctggtcg cgatggtgga cacggaaagc ccactctgcc 240
ccctctcccc	actcgaggcc ggcgatctag agagcccgtt atctgaagag ttcctgcaag 300
aaatgggaaa	catccaagag atttcgcaat ccatcggcga ggatagttct ggaagctttg 360
gcttttacgga	ataccagtat ttaggaagct gtcctggctc agatggctcg gtcatcacgg 420
acacgctttc	accagcttcg agccccctct cggtgactta tcctgtggtc cccggcagcg 480
tggacgagtc	tcccagtggg gcattgaaca tcgaatgtag aatctgcggg gacaaggcct 540
caggctatca	ttacggagtc cacgcgtgtg aaggctgcaa gggcttcttt cggcgaacga 600
ttcgactcaa	gctggtgtat gacaagtgcg accgcagctg caagatccag aaaaagaaca 660
gaaacaaatg	ccagtattgt cgatttcaca agtgcctttc tgtcgggatg tcacacaacg 720
cgattcgttt	tggacgaatg ccaagatctg agaaagcaaa actgaaagca gaaattctta 780
cctgtgaaca	tgacatagaa gattctgaaa ctgcagatct caaatctctg gccaagagaa 840
tctacgaggc	ctacttgaag aacttcaaca tgaacaaggc caaagcccgg gtcatcctct 900
caggaaaggc	cagtaacaat ccaccttttg tcatacatga tatggagaca ctgtgtatgg 960
ctgagaagac	gctggtggcc aagctggtgg ccaatggcat ccagaacaag gaggcggagg 1020
tccgcatctt	tactgtctgc cagtgcacgt cagtggagac cgtcacggag ctcacggaat 1080
tcgccaaggc	catcccaggc ttcgcaaact tggacctgaa cgatcaagtg acattgctaa 1140
aatacggagt	ttatgaggcc atattcgcca tgctgtcttc tgtgatgaac aaagacggga 1200
tgctggtagc	gtatggaaat gggtttataa ctcgtgaatt cctaaaaagc ctaaggaaac 1260

## NLEE01001W00.ST25.txt

```

cgttctgtga tatcatggaa cccaagtttg attttgccat gaagttcaat gcactggaac 1320
tggatgacag tgatatctcc ctttttgtgg ctgctatcat ttgctgtgga gatcgtcctg 1380
gccttctaaa cgtaggacac attgaaaaaa tgcaggaggg tattgtacat gtgctcagac 1440
tccacctgca gagcaaccac cgggacgata tctttctctt cccaaaactt cttcaaaaaa 1500
tggcagacct ccggcagctg gtgacggagc atgctgcagct ggtgcagatc atcaagaaga 1560
cggagtcgga tgctgcgctg caccgcctac tgcaggagat ctacagggac atgtactgag 1620
ttccttcaga tcagccacac cttttccagg agttctgaag ctgacagcac taaaaggag 1680
acgggggagc agcacgattt tgcacaaata tccaccactt taaccttaga gcttggacag 1740
tctgagctgt aggtaaccgg catattattc catatctttg ttttaaccag tacttctaag 1800
agcatagaac tcaaatgctg ggggaggtgg ctaatctcag gactgggaag 1850

```

```

<210> 12
<211> 1609
<212> DNA
<213> HUMAN

```

```

<400> 12
ttcaagtcct tttcttttaa cggattgata ttttgctaga tagagacaaa atatcagtgt 60
gaattacagc aaacccttat tccatgctgt tatgggtgaa actctgggag attctcctat 120
tgaccagaaa agcgattcct tcaactgata actgtctgca aacatatcac aagaaatgac 180
catggttgac acagagatgc cattctggcc caccaacttt gggatcagct ccgtggatct 240
ctccgtaatg gaagaccact cccactcctt tgatatcaag cccttcacta ctgttgactt 300
ctccagcatt tctactccac attacgaaga cattccattc acaagaacag atccagtggg 360
tgcagattac aagtatgacc tgaaacttca agagtaccaa agtgcaatca aagtggagcc 420
tgcacttcca cttattattt ctgagaagac tcagctctac aataagcctc atgaagagcc 480
ttccaactcc ctcatggcaa ttgaatgtcg tgtctgtgga gataaagctt ctggatttca 540
ctatggagtt catgcttgtg aaggatgcaa gggtttcttc cggagaacaa tcagattgaa 600
gcttatctat gacagatgtg atcttaactg tcggatccac aaaaaaagta gaaataaatg 660
tcagtactgt cggtttcaga aatgccttgc agtggggatg tctcataatg ccatcagggt 720
tgggcggatg ccacaggccg agaaggagaa gctgttggcg gagatctcca gtgatatcga 780
ccagctgaat ccagagtccg ctgacctccg ggccctggca aaacatttgt atgactcata 840
cataaagtcc ttcccgtga ccaaagcaaa ggcgagggcg atcttgacag gaaagacaac 900
agacaaatca ccattcgtaa tctatgacat gaattcctta atgatgggag aagataaaat 960
caagttcaaa cacatcaccc ccctgcagga gcagagcaaa gaggtggcca tccgcatctt 1020
tcagggtctg cagtttcgct ccgtggaggc tgtgcaggag atcacagagt atgccaaaag 1080

```

## NLEE01001W00.ST25.txt

cattcctggt	tttgtaaattc	ttgacttgaa	cgaccaagta	actctcctca	aatatggagt	1140
ccacgagatc	atttacacaa	tgctggcctc	cttgatgaat	aaagatgggg	ttctcatatc	1200
cgagggccaa	ggcttcatga	caagggagtt	tctaaagagc	ctgCGaaagc	cttttggtga	1260
ctttatggag	cccaagtttg	agtttgctgt	gaagttcaat	gcactggaat	tagatgacag	1320
cgacttggca	atattttattg	ctgtcattat	tctcagtggg	gaccgcccag	gtttgctgaa	1380
tgtgaagccc	attgaagaca	ttcaagacaa	cctgctacaa	gccctggagc	tccagctgaa	1440
gctgaaccac	cctgagtcct	cacagctgtt	tgccaagctg	ctccagaaaa	tgacagacct	1500
cagacagatt	gtcacggaac	acgtgcagct	actgcagggtg	atcaagaaga	cggagacaga	1560
catgagtctt	cacccgctcc	tgcaggagat	ctacaaggac	ttgtactag		1609

<210> 13  
 <211> 3301  
 <212> DNA  
 <213> HUMAN

<220>  
 <221> misc\_feature  
 <222> (2966)..(2973)  
 <223> n = a, c, g, t

<400> 13						
gaattctgcg	gagcctgCGg	gacggCGggcg	ggttggccccg	taggcagccg	ggacagtgtt	60
gtacagtgtt	ttgggcatgc	acgtgatact	cacacagtgg	cttctgctca	ccaacagatg	120
aagacagatg	caccaacgag	ggctctggaat	ggctctggagt	ggctctggaaa	gcaggggtcag	180
atacccttg	aaaactgaag	cccgtggagc	aatgatctct	acaggactgc	ttcaaggctg	240
atgggaacca	ccctgtagag	gtccatctgc	gttcagaccc	agacgatgcc	agagctatga	300
ctgggcctgc	aggtgtggcg	ccgaggggag	atcagccatg	gagcagccac	aggaggaagc	360
ccctgaggtc	cggaagagg	aggagaaaga	ggaagtggca	gaggcagaag	gagccccaga	420
gctcaatggg	ggaccacagc	atgcacttcc	ttccagcagc	tacacagacc	tctcccggag	480
ctcctcgcca	ccctcactgc	tggaccaact	gcagatgggc	tgtgacgggg	cctcatgcgg	540
cagcctcaac	atggagtgcc	gggtgtgCGg	ggacaaggca	tcgggcttcc	actacgggtgt	600
tcatgcatgt	gaggggtgca	agggcttctt	ccgtcgtacg	atccgcatga	agctggagta	660
cgagaagtgt	gagcgcagct	gcaagattca	gaagaagaac	cgcaacaagt	gccagtactg	720
ccgcttccag	aagtgccttg	cactgggcat	gtcacacaac	gctatccgtt	ttggtcggat	780
gccggagggt	gagaagagga	agctggtggc	agggctgact	gcaaacgagg	ggagccagta	840
caaccacag	gtggccgacc	tgaaggcctt	ctccaagcac	atctacaatg	cctacctgaa	900
aaacttcaac	atgacaaaa	agaaggccccg	cagcatcctc	accggcaaag	ccagccacac	960

## NLEE01001W00.ST25.txt

ggcgcccttt	gtgatccacg	acatcgagac	attgtggcag	gcagagaagg	ggctggtgtg	1020
gaagcagttg	gtgaatggcc	tgcctcccta	caaggagatc	agcgtgcacg	tcttctaccg	1080
ctgccagtgc	accacagtgg	agaccgtgcg	ggagctcact	gagttcgcca	agagcatccc	1140
cagcttcagc	agcctcttcc	tcaacgacca	ggttaccctt	ctcaagtatg	gcgtgcacga	1200
ggccatcttc	gccatgctgg	cctctatcgt	caacaaggac	gggctgctgg	tagccaacgg	1260
cagtggcttt	gtcaccctg	agttcctgcg	cagcctccgc	aaacccttca	gtgatatcat	1320
tgagcctaag	tttgaatttg	ctgtcaagtt	caacgccctg	gaacttgatg	acagtgcact	1380
ggccctattc	attgcggcca	tcattctgtg	tggagaccgg	ccaggcctca	tgaacgttcc	1440
acgggtggag	gctatccagg	acaccatcct	gcgtgccctc	gaattccacc	tgcaggccaa	1500
ccaccctgat	gccagtagc	tcttcccca	gctgctgcag	aagatggctg	acctgcggca	1560
actggtcacc	gagcacgccc	agatgatgca	gcggatcaag	aagaccgaaa	ccgagacctc	1620
gctgcaccct	ctgctccagg	agatctacaa	ggacatgtac	taacggcggc	accaggcct	1680
ccctgcagac	tccaatgggg	ccagcactgg	agggggccac	ccacatgact	tttcattga	1740
ccagctctct	tcctgtcttt	gttgtctccc	tctttctcag	ttcctctttc	ttttctaatt	1800
cctgttgctc	tgtttcttcc	tttctgtagg	tttctctctt	cccttctccc	ttctcccttg	1860
ccctcccttt	ctctctccta	tccccacgtc	tgtcctcctt	tcttattctg	tgagatgttt	1920
tgtattatth	caccagcagc	atagaacagg	acctctgctt	ttgcacacct	tttccccagg	1980
agcagaagag	agtgggcctg	ccctctgccc	catcattgca	cctgcaggct	taggtcctca	2040
cttctgtctc	ctgtcttcag	agcaaaagac	ttgagccatc	caaagaaaca	ctaagctctc	2100
tgggcctggg	ttccaggga	ggctaagcat	ggcctggact	gactgcagcc	ccctatagtc	2160
atggggtccc	tgctgcaaag	gacagtggca	gaccccgga	gtagagccga	gatgcctccc	2220
caagactgtc	attgcccctc	cgatcgtgag	gccaccact	gacccaatga	tcctctccag	2280
cagcacacct	cagccccact	gacaccagc	gtccttccat	cttcacactg	gtttgccagg	2340
ccaatgttgc	tgatggcccc	tccagcacac	acacataagc	actgaaatca	ctttacctgc	2400
aggcaccatg	cacctccctt	ccctccctga	ggcagggtgag	aaccagaga	gaggggcctg	2460
cagggtgagca	ggcagggtg	ggccaggctc	ccggggaggc	aggggtcctg	cagggtcctg	2520
tgggtcagcc	cagcacctcg	cccagtggga	gcttcccggg	ataaactgag	cctgttcatt	2580
ctgatgtcca	tttgtcccaa	tagctctact	gccctcccct	tcccctttac	tcagcccagc	2640
tggccacctc	gaagtctccc	tgcacagcct	ctagtgtccg	gggaccttgt	gggaccagtc	2700
ccacaccgct	ggtccctgcc	ctccccctg	cccagggtga	ggtgcgctca	cctcagagca	2760
gggccaaagc	acagctgggc	atgccatgtc	tgagcggcgc	agagccctcc	aggcctgcag	2820

NLEE01001W00.ST25.txt

gggcaagggg	ctggctggag	tctcagagca	cagaggtagg	agaactggg	ttcaagccca	2880
ggcttccttg	gtcctgcctg	gtcctccctc	ccaaggagcc	attctatgtg	actctgggtg	2940
gaagtgcccc	gcccctgcct	gacggnnnnn	nngatcactc	tctgctggca	ggattcttcc	3000
cgctccccac	ctaccagct	gatgggggtt	ggggtgcttc	tttcagccaa	ggctatgaag	3060
ggacagctgc	tgggaccac	ctccccctt	ccccggccac	atgccgcgtc	cctgccccca	3120
cccgggtctg	gtgctgagga	tacagctctt	ctcagtgtct	gaacaatctc	caaaattgaa	3180
atgtatatatt	ttgctaggag	ccccagcttc	ctgtgttttt	aatataaata	gtgtacacag	3240
actgacgaaa	ctttaataa	atgggaatta	aatatttaaa	aaaaaaagcg	gccgcgaatt	3300
c						3301

<210> 14  
 <211> 3083  
 <212> DNA  
 <213> HUMAN

<400> 14	
aaaaactgca	gccaaacttcc gaggcagcct cattgcccag cggaccccag cctctgccag 60
gttcggtccg	ccatcctcgt cccgtcctcc gccggcccct gccccgcgcc cagggatcct 120
ccagctcctt	tcgcccgcgc cctccgttcg ctccggacac catggacaag ttttggtggc 180
acgcagcctg	gggactctgc ctcgtgccgc tgagcctggc gcagatcgat ttgaatataa 240
cctgccgctt	tgcaggtgta ttceacgtgg agaaaaatgg tcgctacagc atctctcgga 300
cggaggccgc	tgacctctgc aaggctttca atagcacctt gccacaatg gccagatgg 360
agaaagctct	gagcatcgga tttgagacct gcaggatagg gttcatagaa gggcacgtgg 420
tgattccccg	gatccacccc aactccatct gtgcagcaaa caacacaggg gtgtacatcc 480
tcacatccaa	cacctcccag tatgacacat attgcttcaa tgcttcagct ccacctgaag 540
aagattgtac	atcagtcaca gacctgccc atgcctttga tggaccaatt accataacta 600
ttgttaaccg	tgatggcacc cgctatgtcc agaaaggaga atacagaacg aatcctgaag 660
acatctaccc	cagcaaccct actgatgatg acgtgagcag cggcttttct actgtacacc 720
ccatcccaga	cgaagacagt ccctggatca cctcctccag tgaaaggagc agcacttcag 780
gaggttacat	cttttacacc gacagcacag acagaatccc tgctaccact ttgatgagca 840
ctagtgtctac	agcaactgag acagcaacca agaggcaaga aacctgggat tggttttcat 900
ggttgtttct	accatcagag tcaaagaatc atcttcacac aacaacacaa atggctggta 960
cgtcttcaaa	taccatctca gcaggctggg agccaaatga agaaaatgaa gatgaaagag 1020
acagacacct	cagtttttct ggatcaggca ttgatgatga tgaagatttt atctccagca 1080
ccatttcaac	cacaccacgg gcttttgacc acacaaaaca gaaccaggac tggaccagat 1140

NLEE01001W00.ST25.txt

ggaacccaag	ccattcaa	at	ccggaagtgc	tacttcagac	aaccacaagg	atgactgatg	1200
tagacagaaa	tggcaccact	gcttatgaag	gaaactggaa	cccagaagca	caccctcccc		1260
tcattcacca	tgagcatcat	gaggaagaag	agaccccaca	ttctacaagc	acaatccagg		1320
caactcctag	tagtacaacg	gaagaaacag	ctaccagaa	ggaacagtgg	tttggcaaca		1380
gatggcatga	gggatatcgc	caaacacca	aagaagactc	ccatttcaac	ccaatctcac		1440
accccatggg	acgaggtcat	caagcaggaa	gatcgacaac	aggacagct	gcagcctcag		1500
ctcataccag	ccatccaatg	caaggaagga	caacaccaag	cccagaggac	agttcctgga		1560
ctgatttcag	gatggatatg	gactccagtc	atagtataac	gcttcagcct	actgcaaate		1620
caaacacagg	tttgggtggaa	gatttggaca	ggacaggacc	tctttcaatg	acaacgcagc		1680
agagtaattc	tcagagcttc	tctacatcac	atgaaggctt	ggaagaagat	aaagaccatc		1740
caacaacttc	tactctgaca	tcaagcaata	ggaatgatgt	cacaggtgga	agaagagacc		1800
caaatcattc	tgaaggctca	actactttac	tgggaaggta	tacctctcat	taccacaca		1860
cgaaggaaa	caggaccttc	atcccagtga	cctcagctaa	gactgtcaat	cgttccttat		1920
caggagacca	agacacattc	caccccagtg	gggggtcctt	tggagttact	gcagttactg		1980
ttggagattc	caactcta	at	gggtcccata	ccactcatgg	atctgaatca	gatggacact	2040
cacatgggag	tcaagaaggt	ggagcaaaca	caacctctgg	tcctataagg	acaccccaaa		2100
ttccagaatg	gctgatcatc	ttggcatccc	tcttggcctt	ggctttgatt	cttgcagttt		2160
gcattgcagt	caacagtcga	agaagggtgtg	ggcagaagaa	aaagctagt	atcaacagt		2220
gcaatggagc	tgtggaggac	agaaagccaa	gtggactcaa	cggagaggcc	agcaagtctc		2280
aggaaatgg	gcatttggtg	aacaaggagt	cgctcagaac	tccagaccag	tttatgacag		2340
ctgatgagac	aaggaacctg	cagaatgtgg	acatgaagat	tgggggtgta	cacctacacc		2400
attatcttgg	aaagaaacaa	ccgttggaaa	cataaccatt	acaggagagct	gggacactta		2460
acagatgcaa	tgtgctactg	attgtttcat	tgcgaatctt	ttttagcata	aaattttcta		2520
ctctttttgt	tttttgtgtt	ttgttcttta	aagtcagggtc	caatttgtaa	aaacagcatt		2580
gctttctgaa	attagggccc	aattaataat	cagcaagaat	ttgatcgttc	cagttcccac		2640
ttggaggcct	ttcatccctc	gggtgtgcta	tggatggctt	ctaacaaaaa	ctacacatat		2700
gtattcctga	tcgccaacct	ttccccacc	agctaaggac	atttcccagg	gttaataggg		2760
cctggtccct	gggaggaaat	ttgaatgggt	ccattttgcc	cttccatagc	ctaaccctg		2820
ggcattgctt	tccactgagg	ttgggggttg	gggtgtacta	gttacacatc	ttcaacagac		2880
cccctctaga	aatttttcag	atgcttctgg	gagacacca	aagggtgaag	ctatttatct		2940
gtagtaaa	act	atttatctgt	gtttttgaaa	tattaaaccc	tggatcagtc	ctttgatcag	3000
tataattttt	taaagttact	ttgtcagagg	cacaaaagg	tttaaactga	ttcataataa		3060

NLEE01001W00.ST25.txt

atatctgtac ttcttcgatc ttc

3083

<210> 15  
 <211> 2539  
 <212> DNA  
 <213> HUMAN

<400> 15  
 ggagtctctt gctctggttc ttgctgttcc tgctcctgct cccgccgctc cccgtcctgc 60  
 tcgcggaccc aggggcgccc acgccagtga atccctgttg ttactatcca tgccagcacc 120  
 agggcatctg tgtccgcttc ggccttgacc gctaccagtg tgactgcacc cgcacgggct 180  
 attccggccc caactgcacc atccctggcc tgtggacctg gctccggaat tcaactgcggc 240  
 ccagcccctc tttcacccac ttcttgcctc ctcacgggcg ctggttcttg gagtttgtca 300  
 atgccacctt catccgagag atgctcatgc gcctgggtact cacagtgcgc tccaacctta 360  
 tccccagtcc cccacctac aactcagcac atgactacat cagctgggag tctttctcca 420  
 acgtgagcta ttacactcgt attctgccct ctgtgcctaa agattgcccc acacccatgg 480  
 gaaccaaagg gaagaagcag ttgccagatg cccagctcct ggcccgcgcg ttcctgctca 540  
 ggaggaagt catacctgac ccccaaggca ccaacctcat gtttgccttc tttgcacaac 600  
 acttcaccca ccagttcttc aaaacttctg gcaagatggg tcctggcttc accaaggcct 660  
 tggggccatgg ggtagacctc ggccacattt atggagacaa tctggagcgt cagtatcaac 720  
 tgcggctctt taaggatggg aaactcaagt accaggtgct ggatggagaa atgtaccgcg 780  
 cctcggtaga agaggcgcct gtgttgatgc actacccccg aggcattccc cccagagacc 840  
 agatggctgt gggccaggag gtgtttgggc tgcttccttg gctcatgctg tatgccacgc 900  
 tctggctacg tgagcacaac cgtgtgtgtg acctgctgaa ggctgagcac cccacctggg 960  
 gcgatgagca gcttttccag acgaccgcg tcatacctcat aggggagacc atcaagattg 1020  
 tcatacagga gtacgtgcag cagctgagtg gctatttcct gcagctgaaa ttgacccag 1080  
 agctgctgtt cggtgtccag ttccaatacc gcaaccgcat tgccatggag ttcaaccatc 1140  
 tctaccactg gcacccctc atgcctgact cttcaagggt gggctcccag gactacagct 1200  
 acgagcagtt cttgttcaac acctccatgt tgggtgacta tgggggttgag gccctgggtg 1260  
 atgcccttctc tcgccagatt gctggccgga tcggtggggg caggaacatg gaccaccaca 1320  
 tcctgcatgt ggctgtggat gtcatacagg agtctcggga gatgcggctg cagcccttca 1380  
 atgagtaccg caagagggtt ggcatgaaac cctacacctc cttccaggag ctctgtaggag 1440  
 agaaggagat ggcagcagag ttggaggaat tgtatggaga cattgatgctg ttggagttct 1500  
 accctggact gcttcttgaa aagtgccatc caaactctat ctttggggag agtatgatag 1560  
 agattggggc tcccttttcc ctcaagggtc tcctagggaa tcccatctgt tctccggagt 1620

## NLEE01001W00.ST25.txt

```

actggaagcc gagcacattt ggcggcgagg tgggctttaa cattgtcaag acggccacac 1680
tgaagaagct ggtctgcctc aacaccaaga cctgtcccta cgtttccttc cgtgtgccgg 1740
atgccagtca ggatgatggg cctgctgtgg agcgaccatc cacagagctc tgaggggcag 1800
gaaagcagca ttctggaggg gagagctttg tgcttgtcat tccagagtgc tgaggccagg 1860
gctgatggtc ttaaagtctc attttctggg ttggcatggg gagtgttggg gttgacattt 1920
agaactttta gtctcaccca ttatctggaa tattgtgatt ctgtttattc ttccagaatg 1980
ctgaactcct tgtagccct tcagattgtt aggagtgggt ctcatctggg ctgccagaat 2040
actgggttct tagttgacaa cctagaatgt cagatttctg gttgatttgt aacacagtca 2100
ttctaggatg tggagctact gatgaaatct gctagaaagt taggggggttc ttattttgca 2160
ttccagaatc ttgactttct gattgggtgat tcaaagtgtt gtgttcctgg ctgatgatcc 2220
agaacagtgg ctctatccc aaatctgtca gcatctggct gtctagaatg tggatttgat 2280
tcattttcct gttcagttag atatcataga gacggagatc ctaagggtcca acaagaatgc 2340
attccctgaa tctgtgcctg cactgagagg gcaaggaagt ggggtgttct tcttgggacc 2400
cccactaaga ccctgggtctg aggatgtaga gagaacagggt gggctgtatt cacgccattg 2460
gttggaagct accagagctc tatccccatc cagggtcttga ctcatggcag ctgtttctca 2520
tgaagctaataaaaattcgc 2539

```

<210> 16  
<211> 369  
<212> DNA  
<213> HUMAN

```

<400> 16
atgaagcttc tcacgggcct ggttttctgc tccttgggtcc tgggtgtcag cagccgaagc 60
ttcttttcgt tccttggcga ggcttttgat ggggctcggg acatgtggag agcctactct 120
gacatgagag aagccaatta catcgggtca gacaaatact tccatgctcg ggggaactat 180
gatgctgcca aaaggggacc tgggggtgtc tgggctgcag aagcgatcag cgatgccaga 240
gagaatatcc agagattctt tggccatggg gcggaggact cgctggctga tcaggctgcc 300
aatgaatggg gcaggagtgg caaagacccc aatcacttcc gacctgctgg cctgcctgag 360
aaatactga 369

```

<210> 17  
<211> 67  
<212> PRT  
<213> HUMAN

```

<400> 17
Met Thr Ser Lys Leu Ala Val Ala Leu Leu Ala Ala Phe Leu Ile Ser

```

21

NLEE01001w00.ST25.txt

145		150		155		160
Asn	Glu	Ile	Val	Glu	Lys	Leu
				165		
				Leu	Leu	Leu
				Arg	Arg	Lys
				170		
				Phe	Ile	Pro
						175
Pro	Gln	Gly	Ser	Asn	Met	Met
			180			
				Phe	Ala	Phe
				185		
				Phe	Ala	Gln
						190
				His	Phe	Thr
His	Gln	Phe	Phe	Lys	Thr	Asp
		195				
				His	Lys	Arg
				200		
				Gly	Pro	Ala
						205
				Phe	Thr	Asn
Gly	Leu	Gly	His	Gly	Val	Asp
	210					215
				Leu	Asn	His
						Ile
						220
				Tyr	Gly	Glu
						Thr
						Leu
Ala	Arg	Gln	Arg	Lys	Leu	Arg
					230	
				Leu	Phe	Lys
						235
				Asp	Gly	Lys
						Met
						Lys
						240
Gln	Ile	Ile	Asp	Gly	Glu	Met
				245		
				Tyr	Pro	Pro
					250	
				Thr	Val	Lys
						255
				Asp	Thr	Gln
Ala	Glu	Met	Ile	Tyr	Pro	Pro
			260			
				Gln	Val	Pro
				265		
				Pro	Glu	His
						Leu
						270
				Arg	Phe	Ala
Val	Gly	Gln	Glu	Val	Phe	Gly
		275				Leu
						280
				Val	Pro	Gly
						Leu
						285
				Met	Met	Tyr
						Ala
Thr	Ile	Trp	Leu	Arg	Glu	His
	290					295
				Asn	Arg	Val
						Cys
						300
				Asp	Val	Leu
						Lys
						Gln
Glu	His	Pro	Glu	Trp	Gly	Asp
					310	
				Glu	Gln	Leu
						Phe
						315
				Gln	Thr	Ser
						Arg
						Leu
						320
Ile	Leu	Ile	Gly	Glu	Thr	Ile
				325		
				Lys	Ile	Val
						330
				Ile	Glu	Asp
						Tyr
						Val
						335
His	Leu	Ser	Gly	Tyr	His	Phe
			340			
				Lys	Leu	Lys
				345		
				Phe	Asp	Pro
						Glu
						350
				Leu	Leu	Leu
Phe	Asn	Lys	Gln	Phe	Gln	Tyr
		355				
				Gln	Asn	Arg
				360		
				Ile	Ala	Ala
						365
				Glu	Phe	Asn
Thr	Leu	Tyr	His	Trp	His	Pro
	370					375
				Leu	Leu	Pro
						Asp
						380
				Thr	Phe	Gln
						Ile
						His
Asp	Gln	Lys	Tyr	Asn	Tyr	Gln
					390	
				Gln	Gln	Phe
						Ile
						395
				Tyr	Asn	Asn
						Ser
						Ile
						Leu
						400

NLEE01001W00.ST25.txt

Leu Glu His Gly Ile Thr Gln Phe Val Glu Ser Phe Thr Arg Gln Ile  
 405 410 415

Ala Gly Arg Val Ala Gly Gly Arg Asn Val Pro Pro Ala Val Gln Lys  
 420 425 430

Val Ser Gln Ala Ser Ile Asp Gln Ser Arg Gln Met Lys Tyr Gln Ser  
 435 440 445

Phe Asn Glu Tyr Arg Lys Arg Phe Met Leu Lys Pro Tyr Glu Ser Phe  
 450 455 460

Glu Glu Leu Thr Gly Glu Lys Glu Met Ser Ala Glu Leu Glu Ala Leu  
 465 470 475 480

Tyr Gly Asp Ile Asp Ala Val Glu Leu Tyr Pro Ala Leu Leu Val Glu  
 485 490 495

Lys Pro Arg Pro Asp Ala Ile Phe Gly Glu Thr Met Val Glu Val Gly  
 500 505 510

Ala Pro Phe Ser Leu Lys Gly Leu Met Gly Asn Val Ile Cys Ser Pro  
 515 520 525

Ala Tyr Trp Lys Pro Ser Thr Phe Gly Gly Glu Val Gly Phe Gln Ile  
 530 535 540

Ile Asn Thr Ala Ser Ile Gln Ser Leu Ile Cys Asn Asn Val Lys Gly  
 545 550 555 560

Cys Pro Phe Thr Ser Phe Ser Val Pro Asp Pro Glu Leu Ile Lys Thr  
 565 570 575

Val Thr Ile Asn Ala Ser Ser Ser Arg Ser Gly Leu Asp Asp Ile Asn  
 580 585 590

Pro Thr Val Leu Leu Lys Glu Arg Ser Thr Glu Leu  
 595 600

<210> 19  
 <211> 360  
 <212> PRT  
 <213> HUMAN

<400> 19

Met Glu Asp Phe Asn Met Glu Ser Asp Ser Phe Glu Asp Phe Trp Lys  
 1 5 10 15

NLEE01001wo0.ST25.txt

Gly Glu Asp Leu Ser Asn Tyr Ser Tyr Ser Ser Thr Leu Pro Pro Phe  
 20 25 30

Leu Leu Asp Ala Ala Pro Cys Glu Pro Glu Ser Leu Glu Ile Asn Lys  
 35 40 45

Tyr Phe Val Val Ile Ile Tyr Ala Leu Val Phe Leu Leu Ser Leu Leu  
 50 55 60

Gly Asn Ser Leu Val Met Leu Val Ile Leu Tyr Ser Arg Val Gly Arg  
 65 70 75 80

Ser Val Thr Asp Val Tyr Leu Leu Asn Leu Ala Leu Ala Asp Leu Leu  
 85 90 95

Phe Ala Leu Thr Leu Pro Ile Trp Ala Ala Ser Lys Val Asn Gly Trp  
 100 105 110

Ile Phe Gly Thr Phe Leu Cys Lys Val Val Ser Leu Leu Lys Glu Val  
 115 120 125

Asn Phe Tyr Ser Gly Ile Leu Leu Leu Ala Cys Ile Ser Val Asp Arg  
 130 135 140

Tyr Leu Ala Ile Val His Ala Thr Arg Thr Leu Thr Gln Lys Arg Tyr  
 145 150 155 160

Leu Val Lys Phe Ile Cys Leu Ser Ile Trp Gly Leu Ser Leu Leu Leu  
 165 170 175

Ala Leu Pro Val Leu Leu Phe Arg Arg Thr Val Tyr Ser Ser Asn Val  
 180 185 190

Ser Pro Ala Cys Tyr Glu Asp Met Gly Asn Asn Thr Ala Asn Trp Arg  
 195 200 205

Met Leu Leu Arg Ile Leu Pro Gln Ser Phe Gly Phe Ile Val Pro Leu  
 210 215 220

Leu Ile Met Leu Phe Cys Tyr Gly Phe Thr Leu Arg Thr Leu Phe Lys  
 225 230 235 240

Ala His Met Gly Gln Lys His Arg Ala Met Arg Val Ile Phe Ala Val  
 245 250 255

Val Leu Ile Phe Leu Leu Cys Trp Leu Pro Tyr Asn Leu Val Leu Leu  
 260 265 270

NLEE01001W00.ST25.txt

Ala Asp Thr Leu Met Arg Thr Gln Val Ile Gln Glu Thr Cys Glu Arg  
 275 280 285

Arg Asn His Ile Asp Arg Ala Leu Asp Ala Thr Glu Ile Leu Gly Ile  
 290 295 300

Leu His Ser Cys Leu Asn Pro Leu Ile Tyr Ala Phe Ile Gly Gln Lys  
 305 310 315 320

Phe Arg His Gly Leu Leu Lys Ile Leu Ala Ile His Gly Leu Ile Ser  
 325 330 335

Lys Asp Ser Leu Pro Lys Asp Ser Arg Pro Ser Phe Val Gly Ser Ser  
 340 345 350

Ser Gly His Thr Ser Thr Thr Leu  
 355 360

<210> 20  
 <211> 554  
 <212> PRT  
 <213> HUMAN

<400> 20

Met Thr Ala Pro Gly Ala Ala Gly Arg Cys Pro Pro Thr Thr Trp Leu  
 1 5 10 15

Gly Ser Leu Leu Leu Val Cys Leu Leu Ala Ser Arg Ser Ile Thr  
 20 25 30

Glu Glu Val Ser Glu Tyr Cys Ser His Met Ile Gly Ser Gly His Leu  
 35 40 45

Gln Ser Leu Gln Arg Leu Ile Asp Ser Gln Met Glu Thr Ser Cys Gln  
 50 55 60

Ile Thr Phe Glu Phe Val Asp Gln Glu Gln Leu Lys Asp Pro Val Cys  
 65 70 75 80

Tyr Leu Lys Lys Ala Phe Leu Leu Val Gln Asp Ile Met Glu Asp Thr  
 85 90 95

Met Arg Phe Arg Asp Asn Thr Ala Asn Pro Ile Ala Ile Val Gln Leu  
 100 105 110

Gln Glu Leu Ser Leu Arg Leu Lys Ser Cys Phe Thr Lys Asp Tyr Glu  
 115 120 125

## NLEE01001WO0.ST25.txt

Glu His Asp Lys Ala Cys Val Arg Thr Phe Tyr Glu Thr Pro Leu Gln  
 130 135 140  
 Leu Leu Glu Lys Val Lys Asn Val Phe Asn Glu Thr Lys Asn Leu Leu  
 145 150 155 160  
 Asp Lys Asp Trp Asn Ile Phe Ser Lys Asn Cys Asn Asn Ser Phe Ala  
 165 170 175  
 Glu Cys Ser Ser Gln Asp Val Val Thr Lys Pro Asp Cys Asn Cys Leu  
 180 185 190  
 Tyr Pro Lys Ala Ile Pro Ser Ser Asp Pro Ala Ser Val Ser Pro His  
 195 200 205  
 Gln Pro Leu Ala Pro Ser Met Ala Pro Val Ala Gly Leu Thr Trp Glu  
 210 215 220  
 Asp Ser Glu Gly Thr Glu Gly Ser Ser Leu Leu Pro Gly Glu Gln Pro  
 225 230 235 240  
 Leu His Thr Val Asp Pro Gly Ser Ala Lys Gln Arg Pro Pro Arg Ser  
 245 250 255  
 Thr Cys Gln Ser Phe Glu Pro Pro Glu Thr Pro Val Val Lys Asp Ser  
 260 265 270  
 Thr Ile Gly Gly Ser Pro Gln Pro Arg Pro Ser Val Gly Ala Phe Asn  
 275 280 285  
 Pro Gly Met Glu Asp Ile Leu Asp Ser Ala Met Gly Thr Asn Trp Val  
 290 295 300  
 Pro Glu Glu Ala Ser Gly Glu Ala Ser Glu Ile Pro Val Pro Gln Gly  
 305 310 315 320  
 Thr Glu Leu Ser Pro Ser Arg Pro Gly Gly Gly Ser Met Gln Thr Glu  
 325 330 335  
 Pro Ala Arg Pro Ser Asn Phe Leu Ser Ala Ser Ser Pro Leu Pro Ala  
 340 345 350  
 Ser Ala Lys Gly Gln Gln Pro Ala Asp Val Thr Ala Thr Ala Leu Pro  
 355 360 365  
 Arg Val Gly Pro Val Met Pro Thr Gly Gln Asp Trp Asn His Thr Pro  
 370 375 380

NLEE01001wo0.ST25.txt

Gln Lys Thr Asp His Pro Ser Ala Leu Leu Arg Asp Pro Pro Glu Pro  
 385 390 395 400

Gly Ser Pro Arg Ile Ser Ser Leu Arg Pro Gln Ala Leu Ser Asn Pro  
 405 410 415

Ser Thr Leu Ser Ala Gln Pro Gln Leu Ser Arg Ser His Ser Ser Gly  
 420 425 430

Ser Val Leu Pro Leu Gly Glu Leu Glu Gly Arg Arg Ser Thr Arg Asp  
 435 440 445

Arg Thr Ser Pro Ala Glu Pro Glu Ala Ala Pro Ala Ser Glu Gly Ala  
 450 455 460

Ala Arg Pro Leu Pro Arg Phe Asn Ser Val Pro Leu Thr Asp Thr Gly  
 465 470 475 480

His Glu Arg Gln Ser Glu Gly Ser Ser Ser Pro Gln Leu Gln Glu Ser  
 485 490 495

Val Phe His Leu Leu Val Pro Ser Val Ile Leu Val Leu Leu Ala Val  
 500 505 510

Gly Gly Leu Leu Phe Tyr Arg Trp Arg Arg Arg Ser His Gln Glu Pro  
 515 520 525

Gln Arg Ala Asp Ser Pro Leu Glu Gln Pro Glu Gly Ser Pro Leu Thr  
 530 535 540

Gln Asp Asp Arg Gln Val Glu Leu Pro Val  
 545 550

<210> 21  
 <211> 107  
 <212> PRT  
 <213> HUMAN

<400> 21

Met Ala Arg Ala Ala Leu Ser Ala Ala Pro Ser Asn Pro Arg Leu Leu  
 1 5 10 15

Arg Val Ala Leu Leu Leu Leu Leu Val Ala Ala Gly Arg Arg Ala  
 20 25 30

Ala Gly Ala Ser Val Ala Thr Glu Leu Arg Cys Gln Cys Leu Gln Thr  
 35 40 45

NLEE01001W00.ST25.txt

Leu Gln Gly Ile His Pro Lys Asn Ile Gln Ser Val Asn Val Lys Ser  
 50 55 60

Pro Gly Pro His Cys Ala Gln Thr Glu Val Ile Ala Thr Leu Lys Asn  
 65 70 75 80

Gly Arg Lys Ala Cys Leu Asn Pro Ala Ser Pro Ile Val Lys Lys Ile  
 85 90 95

Ile Glu Lys Met Leu Asn Ser Asp Lys Ser Asn  
 100 105

<210> 22  
 <211> 106  
 <212> PRT  
 <213> HUMAN

<400> 22

Met Ala His Ala Thr Leu Ser Ala Ala Pro Ser Asn Pro Arg Leu Leu  
 1 5 10 15

Arg Val Ala Leu Leu Leu Leu Leu Val Gly Ser Arg Arg Ala Ala  
 20 25 30

Gly Ala Ser Val Val Thr Glu Leu Arg Cys Gln Cys Leu Gln Thr Leu  
 35 40 45

Gln Gly Ile His Leu Lys Asn Ile Gln Ser Val Asn Val Arg Ser Pro  
 50 55 60

Gly Pro His Cys Ala Gln Thr Glu Val Ile Ala Thr Leu Lys Asn Gly  
 65 70 75 80

Lys Lys Ala Cys Leu Asn Pro Ala Ser Pro Met Val Gln Lys Ile Ile  
 85 90 95

Glu Lys Ile Leu Asn Lys Gly Ser Thr Asn  
 100 105

<210> 23  
 <211> 300  
 <212> PRT  
 <213> HUMAN

<400> 23

Met Arg Ile Ala Val Ile Cys Phe Cys Leu Leu Gly Ile Thr Cys Ala  
 1 5 10 15

## NLEE01001W00.ST25.txt

Ile Pro Val Lys Gln Ala Asp Ser Gly Ser Ser Glu Glu Lys Gln Leu  
 20 25 30  
 Tyr Asn Lys Tyr Pro Asp Ala Val Ala Thr Trp Leu Asn Pro Asp Pro  
 35 40 45  
 Ser Gln Lys Gln Asn Leu Leu Ala Pro Gln Thr Leu Pro Ser Lys Ser  
 50 55 60  
 Asn Glu Ser His Asp His Met Asp Asp Met Asp Asp Glu Asp Asp Asp  
 65 70 75 80  
 Asp His Val Asp Ser Gln Asp Ser Ile Asp Ser Asn Asp Ser Asp Asp  
 85 90 95  
 Val Asp Asp Thr Asp Asp Ser His Gln Ser Asp Glu Ser His His Ser  
 100 105 110  
 Asp Glu Ser Asp Glu Leu Val Thr Asp Phe Pro Thr Asp Leu Pro Ala  
 115 120 125  
 Thr Glu Val Phe Thr Pro Val Val Pro Thr Val Asp Thr Tyr Asp Gly  
 130 135 140  
 Arg Gly Asp Ser Val Val Tyr Gly Leu Arg Ser Lys Ser Lys Lys Phe  
 145 150 155 160  
 Arg Arg Pro Asp Ile Gln Tyr Pro Asp Ala Thr Asp Glu Asp Ile Thr  
 165 170 175  
 Ser His Met Glu Ser Glu Glu Leu Asn Gly Ala Tyr Lys Ala Ile Pro  
 180 185 190  
 Val Ala Gln Asp Leu Asn Ala Pro Ser Asp Trp Asp Ser Arg Gly Lys  
 195 200 205  
 Asp Ser Tyr Glu Thr Ser Gln Leu Asp Asp Gln Ser Ala Glu Thr His  
 210 215 220  
 Ser His Lys Gln Ser Arg Leu Tyr Lys Arg Lys Ala Asn Asp Glu Ser  
 225 230 235 240  
 Asn Glu His Ser Asp Val Ile Asp Ser Gln Glu Leu Ser Lys Val Ser  
 245 250 255  
 Arg Glu Phe His Ser His Glu Phe His Ser His Glu Asp Met Leu Val  
 260 265 270

NLEE01001w00.ST25.txt

Val Asp Pro Lys Ser Lys Glu Glu Asp Lys His Leu Lys Phe Arg Ile  
 275 280 285

Ser His Glu Leu Asp Ser Ala Ser Ser Glu Val Asn  
 290 295 300

<210> 24  
 <211> 295  
 <212> PRT  
 <213> HUMAN

<400> 24

Met Glu His Gln Leu Leu Cys Cys Glu Val Glu Thr Ile Arg Arg Ala  
 1 5 10 15

Tyr Pro Asp Ala Asn Leu Leu Asn Asp Arg Val Leu Arg Ala Met Leu  
 20 25 30

Lys Ala Glu Glu Thr Cys Ala Pro Ser Val Ser Tyr Phe Lys Cys Val  
 35 40 45

Gln Lys Glu Val Leu Pro Ser Met Arg Lys Ile Val Ala Thr Trp Met  
 50 55 60

Leu Glu Val Cys Glu Glu Gln Lys Cys Glu Glu Glu Val Phe Pro Leu  
 65 70 75 80

Ala Met Asn Tyr Leu Asp Arg Phe Leu Ser Leu Glu Pro Val Lys Lys  
 85 90 95

Ser Arg Leu Gln Leu Leu Gly Ala Thr Cys Met Phe Val Ala Ser Lys  
 100 105 110

Met Lys Glu Thr Ile Pro Leu Thr Ala Glu Lys Leu Cys Ile Tyr Thr  
 115 120 125

Asp Gly Ser Ile Arg Pro Glu Glu Leu Leu Gln Met Glu Leu Leu Leu  
 130 135 140

Val Asn Lys Leu Lys Trp Asn Leu Ala Ala Met Thr Pro His Asp Phe  
 145 150 155 160

Ile Glu His Phe Leu Ser Lys Met Pro Glu Ala Glu Glu Asn Lys Gln  
 165 170 175

Ile Ile Arg Lys His Ala Gln Thr Phe Val Ala Ser Cys Ala Thr Asp  
 180 185 190

NLEE01001W00.ST25.txt

Val Lys Phe Ile Ser Asn Pro Pro Ser Met Val Ala Ala Gly Ser Val  
 195 200 205  
 Val Ala Ala Val Gln Gly Leu Asn Leu Arg Ser Pro Asn Asn Phe Leu  
 210 215 220  
 Ser Tyr Tyr Arg Leu Thr Arg Phe Leu Ser Arg Val Ile Lys Cys Asp  
 225 230 235 240  
 Pro Asp Cys Leu Arg Ala Cys Gln Glu Gln Ile Glu Ala Leu Leu Glu  
 245 250 255  
 Ser Ser Leu Arg Gln Ala Gln Gln Asn Met Asp Pro Lys Ala Ala Glu  
 260 265 270  
 Glu Glu Glu Glu Glu Glu Glu Val Asp Leu Ala Cys Thr Pro Thr  
 275 280 285  
 Asp Val Arg Asp Val Asp Ile  
 290 295

<210> 25  
 <211> 439  
 <212> PRT  
 <213> HUMAN

<400> 25

Met Pro Leu Asn Val Ser Phe Thr Asn Arg Asn Tyr Asp Leu Asp Tyr  
 1 5 10 15  
 Asp Ser Val Gln Pro Tyr Phe Tyr Cys Asp Glu Glu Glu Asn Phe Tyr  
 20 25 30  
 Gln Gln Gln Gln Gln Ser Glu Leu Gln Pro Pro Ala Pro Ser Glu Asp  
 35 40 45  
 Ile Trp Lys Lys Phe Glu Leu Leu Pro Thr Pro Pro Leu Ser Pro Ser  
 50 55 60  
 Arg Arg Ser Gly Leu Cys Ser Pro Ser Tyr Val Ala Val Thr Pro Phe  
 65 70 75 80  
 Ser Leu Arg Gly Asp Asn Asp Gly Gly Gly Gly Ser Phe Ser Thr Ala  
 85 90 95  
 Asp Gln Leu Glu Met Val Thr Glu Leu Leu Gly Gly Asp Met Val Asn  
 100 105 110

NLEE01001W00.ST25.txt

Gln Ser Phe Ile Cys Asp Pro Asp Asp Glu Thr Phe Ile Lys Asn Ile  
 115 120 125  
 Ile Ile Gln Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Ala Lys Leu  
 130 135 140  
 Val Ser Glu Lys Leu Ala Ser Tyr Gln Ala Ala Arg Lys Asp Ser Gly  
 145 150 155 160  
 Ser Pro Asn Pro Ala Arg Gly His Ser Val Cys Ser Thr Ser Ser Leu  
 165 170 175  
 Tyr Leu Gln Asp Leu Ser Ala Ala Ala Ser Glu Cys Ile Asp Pro Ser  
 180 185 190  
 Val Val Phe Pro Tyr Pro Leu Asn Asp Ser Ser Ser Pro Lys Ser Cys  
 195 200 205  
 Ala Ser Gln Asp Ser Ser Ala Phe Ser Pro Ser Ser Asp Ser Leu Leu  
 210 215 220  
 Ser Ser Thr Glu Ser Ser Pro Gln Gly Ser Pro Glu Pro Leu Val Leu  
 225 230 235 240  
 His Glu Glu Thr Pro Pro Thr Thr Ser Ser Asp Ser Glu Glu Glu Gln  
 245 250 255  
 Glu Asp Glu Glu Glu Ile Asp Val Val Ser Val Glu Lys Arg Gln Ala  
 260 265 270  
 Pro Gly Lys Arg Ser Glu Ser Gly Ser Pro Ser Ala Gly Gly His Ser  
 275 280 285  
 Lys Pro Pro His Ser Pro Leu Val Leu Lys Arg Cys His Val Ser Thr  
 290 295 300  
 His Gln His Asn Tyr Ala Ala Pro Pro Ser Thr Arg Lys Asp Tyr Pro  
 305 310 315 320  
 Ala Ala Lys Arg Val Lys Leu Asp Ser Val Arg Val Leu Arg Gln Ile  
 325 330 335  
 Ser Asn Asn Arg Lys Cys Thr Ser Pro Arg Ser Ser Asp Thr Glu Glu  
 340 345 350  
 Asn Val Lys Arg Arg Thr His Asn Val Leu Glu Arg Gln Arg Arg Asn

33

NLEE01001W00.ST25.txt

130

135

140

Thr Ser Met Thr Asp Phe Tyr His Ser Lys Arg Arg Leu Ile Phe Ser  
 145 150 155 160

Lys Arg Lys Pro

<210> 27  
 <211> 468  
 <212> PRT  
 <213> HUMAN

<400> 27

Met Val Asp Thr Glu Ser Pro Leu Cys Pro Leu Ser Pro Leu Glu Ala  
 1 5 10 15

Gly Asp Leu Glu Ser Pro Leu Ser Glu Glu Phe Leu Gln Glu Met Gly  
 20 25 30

Asn Ile Gln Glu Ile Ser Gln Ser Ile Gly Glu Asp Ser Ser Gly Ser  
 35 40 45

Phe Gly Phe Thr Glu Tyr Gln Tyr Leu Gly Ser Cys Pro Gly Ser Asp  
 50 55 60

Gly Ser Val Ile Thr Asp Thr Leu Ser Pro Ala Ser Ser Pro Ser Ser  
 65 70 75 80

Val Thr Tyr Pro Val Val Pro Gly Ser Val Asp Glu Ser Pro Ser Gly  
 85 90 95

Ala Leu Asn Ile Glu Cys Arg Ile Cys Gly Asp Lys Ala Ser Gly Tyr  
 100 105 110

His Tyr Gly Val His Ala Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg  
 115 120 125

Thr Ile Arg Leu Lys Leu Val Tyr Asp Lys Cys Asp Arg Ser Cys Lys  
 130 135 140

Ile Gln Lys Lys Asn Arg Asn Lys Cys Gln Tyr Cys Arg Phe His Lys  
 145 150 155 160

Cys Leu Ser Val Gly Met Ser His Asn Ala Ile Arg Phe Gly Arg Met  
 165 170 175

Pro Arg Ser Glu Lys Ala Lys Leu Lys Ala Glu Ile Leu Thr Cys Glu

NLEE01001W00.ST25.txt

180

185

190

His Asp Ile Glu Asp Ser Glu Thr Ala Asp Leu Lys Ser Leu Ala Lys  
 195 200 205

Arg Ile Tyr Glu Ala Tyr Leu Lys Asn Phe Asn Met Asn Lys Val Lys  
 210 215 220

Ala Arg Val Ile Leu Ser Gly Lys Ala Ser Asn Asn Pro Pro Phe Val  
 225 230 235 240

Ile His Asp Met Glu Thr Leu Cys Met Ala Glu Lys Thr Leu Val Ala  
 245 250 255

Lys Leu Val Ala Asn Gly Ile Gln Asn Lys Glu Ala Glu Val Arg Ile  
 260 265 270

Phe His Cys Cys Gln Cys Thr Ser Val Glu Thr Val Thr Glu Leu Thr  
 275 280 285

Glu Phe Ala Lys Ala Ile Pro Gly Phe Ala Asn Leu Asp Leu Asn Asp  
 290 295 300

Gln Val Thr Leu Leu Lys Tyr Gly Val Tyr Glu Ala Ile Phe Ala Met  
 305 310 315 320

Leu Ser Ser Val Met Asn Lys Asp Gly Met Leu Val Ala Tyr Gly Asn  
 325 330 335

Gly Phe Ile Thr Arg Glu Phe Leu Lys Ser Leu Arg Lys Pro Phe Cys  
 340 345 350

Asp Ile Met Glu Pro Lys Phe Asp Phe Ala Met Lys Phe Asn Ala Leu  
 355 360 365

Glu Leu Asp Asp Ser Asp Ile Ser Leu Phe Val Ala Ala Ile Ile Cys  
 370 375 380

Cys Gly Asp Arg Pro Gly Leu Leu Asn Val Gly His Ile Glu Lys Met  
 385 390 395 400

Gln Glu Gly Ile Val His Val Leu Arg Leu His Leu Gln Ser Asn His  
 405 410 415

Pro Asp Asp Ile Phe Leu Phe Pro Lys Leu Leu Gln Lys Met Ala Asp  
 420 425 430

NLEE01001w00.ST25.txt

Leu Arg Gln Leu Val Thr Glu His Ala Gln Leu Val Gln Ile Ile Lys  
 435 440 445

Lys Thr Glu Ser Asp Ala Ala Leu His Pro Leu Leu Gln Glu Ile Tyr  
 450 455 460

Arg Asp Met Tyr  
 465

<210> 28  
 <211> 505  
 <212> PRT  
 <213> HUMAN

<400> 28

Met Gly Glu Thr Leu Gly Asp Ser Pro Ile Asp Pro Glu Ser Asp Ser  
 1 5 10 15

Phe Thr Asp Thr Leu Ser Ala Asn Ile Ser Gln Glu Met Thr Met Val  
 20 25 30

Asp Thr Glu Met Pro Phe Trp Pro Thr Asn Phe Gly Ile Ser Ser Val  
 35 40 45

Asp Leu Ser Val Met Glu Asp His Ser His Ser Phe Asp Ile Lys Pro  
 50 55 60

Phe Thr Thr Val Asp Phe Ser Ser Ile Ser Thr Pro His Tyr Glu Asp  
 65 70 75 80

Ile Pro Phe Thr Arg Thr Asp Pro Val Val Ala Asp Tyr Lys Tyr Asp  
 85 90 95

Leu Lys Leu Gln Glu Tyr Gln Ser Ala Ile Lys Val Glu Pro Ala Ser  
 100 105 110

Pro Pro Tyr Tyr Ser Glu Lys Thr Gln Leu Tyr Asn Lys Pro His Glu  
 115 120 125

Glu Pro Ser Asn Ser Leu Met Ala Ile Glu Cys Arg Val Cys Gly Asp  
 130 135 140

Lys Ala Ser Gly Phe His Tyr Gly Val His Ala Cys Glu Gly Cys Lys  
 145 150 155 160

Gly Phe Phe Arg Arg Thr Ile Arg Leu Lys Leu Ile Tyr Asp Arg Cys  
 165 170 175

NLEE01001W00.ST25.txt

Asp Leu Asn Cys Arg Ile His Lys Lys Ser Arg Asn Lys Cys Gln Tyr  
 180 185 190  
 Cys Arg Phe Gln Lys Cys Leu Ala Val Gly Met Ser His Asn Ala Ile  
 195 200 205  
 Arg Phe Gly Arg Met Pro Gln Ala Glu Lys Glu Lys Leu Leu Ala Glu  
 210 215 220  
 Ile Ser Ser Asp Ile Asp Gln Leu Asn Pro Glu Ser Ala Asp Leu Arg  
 225 230 235 240  
 Ala Leu Ala Lys His Leu Tyr Asp Ser Tyr Ile Lys Ser Phe Pro Leu  
 245 250 255  
 Thr Lys Ala Lys Ala Arg Ala Ile Leu Thr Gly Lys Thr Thr Asp Lys  
 260 265 270  
 Ser Pro Phe Val Ile Tyr Asp Met Asn Ser Leu Met Met Gly Glu Asp  
 275 280 285  
 Lys Ile Lys Phe Lys His Ile Thr Pro Leu Gln Glu Gln Ser Lys Glu  
 290 295 300  
 Val Ala Ile Arg Ile Phe Gln Gly Cys Gln Phe Arg Ser Val Glu Ala  
 305 310 315 320  
 Val Gln Glu Ile Thr Glu Tyr Ala Lys Ser Ile Pro Gly Phe Val Asn  
 325 330 335  
 Leu Asp Leu Asn Asp Gln Val Thr Leu Leu Lys Tyr Gly Val His Glu  
 340 345 350  
 Ile Ile Tyr Thr Met Leu Ala Ser Leu Met Asn Lys Asp Gly Val Leu  
 355 360 365  
 Ile Ser Glu Gly Gln Gly Phe Met Thr Arg Glu Phe Leu Lys Ser Leu  
 370 375 380  
 Arg Lys Pro Phe Gly Asp Phe Met Glu Pro Lys Phe Glu Phe Ala Val  
 385 390 395 400  
 Lys Phe Asn Ala Leu Glu Leu Asp Asp Ser Asp Leu Ala Ile Phe Ile  
 405 410 415  
 Ala Val Ile Ile Leu Ser Gly Asp Arg Pro Gly Leu Leu Asn Val Lys  
 420 425 430

NLEE01001w00.ST25.txt

Pro Ile Glu Asp Ile Gln Asp Asn Leu Leu Gln Ala Leu Glu Leu Gln  
 435 440 445

Leu Lys Leu Asn His Pro Glu Ser Ser Gln Leu Phe Ala Lys Leu Leu  
 450 455 460

Gln Lys Met Thr Asp Leu Arg Gln Ile Val Thr Glu His Val Gln Leu  
 465 470 475 480

Leu Gln Val Ile Lys Lys Thr Glu Thr Asp Met Ser Leu His Pro Leu  
 485 490 495

Leu Gln Glu Ile Tyr Lys Asp Leu Tyr  
 500 505

<210> 29  
 <211> 441  
 <212> PRT  
 <213> HUMAN

<400> 29

Met Glu Gln Pro Gln Glu Glu Ala Pro Glu Val Arg Glu Glu Glu Glu  
 1 5 10 15

Lys Glu Glu Val Ala Glu Ala Glu Gly Ala Pro Glu Leu Asn Gly Gly  
 20 25 30

Pro Gln His Ala Leu Pro Ser Ser Ser Tyr Thr Asp Leu Ser Arg Ser  
 35 40 45

Ser Ser Pro Pro Ser Leu Leu Asp Gln Leu Gln Met Gly Cys Asp Gly  
 50 55 60

Ala Ser Cys Gly Ser Leu Asn Met Glu Cys Arg Val Cys Gly Asp Lys  
 65 70 75 80

Ala Ser Gly Phe His Tyr Gly Val His Ala Cys Glu Gly Cys Lys Gly  
 85 90 95

Phe Phe Arg Arg Thr Ile Arg Met Lys Leu Glu Tyr Glu Lys Cys Glu  
 100 105 110

Arg Ser Cys Lys Ile Gln Lys Lys Asn Arg Asn Lys Cys Gln Tyr Cys  
 115 120 125

Arg Phe Gln Lys Cys Leu Ala Leu Gly Met Ser His Asn Ala Ile Arg  
 130 135 140

## NLEE01001w00.ST25.txt

Phe Gly Arg Met Pro Glu Ala Glu Lys Arg Lys Leu Val Ala Gly Leu  
 145 150 155 160  
 Thr Ala Asn Glu Gly Ser Gln Tyr Asn Pro Gln Val Ala Asp Leu Lys  
 165 170 175  
 Ala Phe Ser Lys His Ile Tyr Asn Ala Tyr Leu Lys Asn Phe Asn Met  
 180 185 190  
 Thr Lys Lys Lys Ala Arg Ser Ile Leu Thr Gly Lys Ala Ser His Thr  
 195 200 205  
 Ala Pro Phe Val Ile His Asp Ile Glu Thr Leu Trp Gln Ala Glu Lys  
 210 215 220  
 Gly Leu Val Trp Lys Gln Leu Val Asn Gly Leu Pro Pro Tyr Lys Glu  
 225 230 235 240  
 Ile Ser Val His Val Phe Tyr Arg Cys Gln Cys Thr Thr Val Glu Thr  
 245 250 255  
 Val Arg Glu Leu Thr Glu Phe Ala Lys Ser Ile Pro Ser Phe Ser Ser  
 260 265 270  
 Leu Phe Leu Asn Asp Gln Val Thr Leu Leu Lys Tyr Gly Val His Glu  
 275 280 285  
 Ala Ile Phe Ala Met Leu Ala Ser Ile Val Asn Lys Asp Gly Leu Leu  
 290 295 300  
 Val Ala Asn Gly Ser Gly Phe Val Thr Arg Glu Phe Leu Arg Ser Leu  
 305 310 315 320  
 Arg Lys Pro Phe Ser Asp Ile Ile Glu Pro Lys Phe Glu Phe Ala Val  
 325 330 335  
 Lys Phe Asn Ala Leu Glu Leu Asp Asp Ser Asp Leu Ala Leu Phe Ile  
 340 345 350  
 Ala Ala Ile Ile Leu Cys Gly Asp Arg Pro Gly Leu Met Asn Val Pro  
 355 360 365  
 Arg Val Glu Ala Ile Gln Asp Thr Ile Leu Arg Ala Leu Glu Phe His  
 370 375 380  
 Leu Gln Ala Asn His Pro Asp Ala Gln Tyr Leu Phe Pro Lys Leu Leu  
 385 390 395 400

NLEE01001W00.ST25.txt

Gln Lys Met Ala Asp Leu Arg Gln Leu Val Thr Glu His Ala Gln Met  
 405 410 415

Met Gln Arg Ile Lys Lys Thr Glu Thr Glu Thr Ser Leu His Pro Leu  
 420 425 430

Leu Gln Glu Ile Tyr Lys Asp Met Tyr  
 435 440

<210> 30  
 <211> 742  
 <212> PRT  
 <213> HUMAN

<400> 30

Met Asp Lys Phe Trp Trp His Ala Ala Trp Gly Leu Cys Leu Val Pro  
 1 5 10 15

Leu Ser Leu Ala Gln Ile Asp Leu Asn Ile Thr Cys Arg Phe Ala Gly  
 20 25 30

Val Phe His Val Glu Lys Asn Gly Arg Tyr Ser Ile Ser Arg Thr Glu  
 35 40 45

Ala Ala Asp Leu Cys Lys Ala Phe Asn Ser Thr Leu Pro Thr Met Ala  
 50 55 60

Gln Met Glu Lys Ala Leu Ser Ile Gly Phe Glu Thr Cys Arg Tyr Gly  
 65 70 75 80

Phe Ile Glu Gly His Val Val Ile Pro Arg Ile His Pro Asn Ser Ile  
 85 90 95

Cys Ala Ala Asn Asn Thr Gly Val Tyr Ile Leu Thr Ser Asn Thr Ser  
 100 105 110

Gln Tyr Asp Thr Tyr Cys Phe Asn Ala Ser Ala Pro Pro Glu Glu Asp  
 115 120 125

Cys Thr Ser Val Thr Asp Leu Pro Asn Ala Phe Asp Gly Pro Ile Thr  
 130 135 140

Ile Thr Ile Val Asn Arg Asp Gly Thr Arg Tyr Val Gln Lys Gly Glu  
 145 150 155 160

Tyr Arg Thr Asn Pro Glu Asp Ile Tyr Pro Ser Asn Pro Thr Asp Asp  
 165 170 175

NLEE01001w00.ST25.txt

Asp Val Ser Ser Gly Ser Ser Ser Glu Arg Ser Ser Thr Ser Gly Gly  
 180 185 190  
 Tyr Ile Phe Tyr Thr Phe Ser Thr Val His Pro Ile Pro Asp Glu Asp  
 195 200 205  
 Ser Pro Trp Ile Thr Asp Ser Thr Asp Arg Ile Pro Ala Thr Thr Leu  
 210 215 220  
 Met Ser Thr Ser Ala Thr Ala Thr Glu Thr Ala Thr Lys Arg Gln Glu  
 225 230 235 240  
 Thr Trp Asp Trp Phe Ser Trp Leu Phe Leu Pro Ser Glu Ser Lys Asn  
 245 250 255  
 His Leu His Thr Thr Thr Gln Met Ala Gly Thr Ser Ser Asn Thr Ile  
 260 265 270  
 Ser Ala Gly Trp Glu Pro Asn Glu Glu Asn Glu Asp Glu Arg Asp Arg  
 275 280 285  
 His Leu Ser Phe Ser Gly Ser Gly Ile Asp Asp Asp Glu Asp Phe Ile  
 290 295 300  
 Ser Ser Thr Ile Ser Thr Thr Pro Arg Ala Phe Asp His Thr Lys Gln  
 305 310 315 320  
 Asn Gln Asp Trp Thr Gln Trp Asn Pro Ser His Ser Asn Pro Glu Val  
 325 330 335  
 Leu Leu Gln Thr Thr Thr Arg Met Thr Asp Val Asp Arg Asn Gly Thr  
 340 345 350  
 Thr Ala Tyr Glu Gly Asn Trp Asn Pro Glu Ala His Pro Pro Leu Ile  
 355 360 365  
 His His Glu His His Glu Glu Glu Glu Thr Pro His Ser Thr Ser Thr  
 370 375 380  
 Ile Gln Ala Thr Pro Ser Ser Thr Thr Glu Glu Thr Ala Thr Gln Lys  
 385 390 395 400  
 Glu Gln Trp Phe Gly Asn Arg Trp His Glu Gly Tyr Arg Gln Thr Pro  
 405 410 415  
 Lys Glu Asp Ser His Ser Thr Thr Gly Thr Ala Ala Ala Ser Ala His  
 41

NLEE01001wo0.ST25.txt

420

425

430

Thr Ser His Pro Met Gln Gly Arg Thr Thr Pro Ser Pro Glu Asp Ser  
 435 440 445

Ser Trp Thr Asp Phe Phe Asn Pro Ile Ser His Pro Met Gly Arg Gly  
 450 455 460

His Gln Ala Gly Arg Arg Met Asp Met Asp Ser Ser His Ser Ile Thr  
 465 470 475 480

Leu Gln Pro Thr Ala Asn Pro Asn Thr Gly Leu Val Glu Asp Leu Asp  
 485 490 495

Arg Thr Gly Pro Leu Ser Met Thr Thr Gln Gln Ser Asn Ser Gln Ser  
 500 505 510

Phe Ser Thr Ser His Glu Gly Leu Glu Glu Asp Lys Asp His Pro Thr  
 515 520 525

Thr Ser Thr Leu Thr Ser Ser Asn Arg Asn Asp Val Thr Gly Gly Arg  
 530 535 540

Arg Asp Pro Asn His Ser Glu Gly Ser Thr Thr Leu Leu Glu Gly Tyr  
 545 550 555 560

Thr Ser His Tyr Pro His Thr Lys Glu Ser Arg Thr Phe Ile Pro Val  
 565 570 575

Thr Ser Ala Lys Thr Gly Ser Phe Gly Val Thr Ala Val Thr Val Gly  
 580 585 590

Asp Ser Asn Ser Asn Val Asn Arg Ser Leu Ser Gly Asp Gln Asp Thr  
 595 600 605

Phe His Pro Ser Gly Gly Ser His Thr Thr His Gly Ser Glu Ser Asp  
 610 615 620

Gly His Ser His Gly Ser Gln Glu Gly Gly Ala Asn Thr Thr Ser Gly  
 625 630 635 640

Pro Ile Arg Thr Pro Gln Ile Pro Glu Trp Leu Ile Ile Leu Ala Ser  
 645 650 655

Leu Leu Ala Leu Ala Leu Ile Leu Ala Val Cys Ile Ala Val Asn Ser  
 660 665 670

NLEE01001W00.ST25.txt

Arg Arg Arg Cys Gly Gln Lys Lys Lys Leu Val Ile Asn Ser Gly Asn  
 675 680 685

Gly Ala Val Glu Asp Arg Lys Pro Ser Gly Leu Asn Gly Glu Ala Ser  
 690 695 700

Lys Ser Gln Glu Met Val His Leu Val Asn Lys Glu Ser Ser Glu Thr  
 705 710 715 720

Pro Asp Gln Phe Met Thr Ala Asp Glu Thr Arg Asn Leu Gln Asn Val  
 725 730 735

Asp Met Lys Ile Gly Val  
 740

<210> 31  
 <211> 489  
 <212> PRT  
 <213> HUMAN

<400> 31

Met Leu Met Arg Leu Val Leu Thr Val Arg Ser Asn Leu Ile Pro Ser  
 1 5 10 15

Pro Pro Thr Tyr Asn Ser Ala His Asp Tyr Ile Ser Trp Glu Ser Phe  
 20 25 30

Ser Asn Val Ser Tyr Tyr Thr Arg Ile Leu Pro Ser Val Pro Lys Asp  
 35 40 45

Cys Pro Thr Pro Met Gly Thr Lys Gly Lys Lys Gln Leu Pro Asp Ala  
 50 55 60

Gln Leu Leu Ala Arg Arg Phe Leu Leu Arg Arg Lys Phe Ile Pro Asp  
 65 70 75 80

Pro Gln Gly Thr Asn Leu Met Phe Ala Phe Phe Ala Gln His Phe Thr  
 85 90 95

His Gln Phe Phe Lys Thr Ser Gly Lys Met Gly Pro Gly Phe Thr Lys  
 100 105 110

Ala Leu Gly His Gly Val Asp Leu Gly His Ile Tyr Gly Asp Asn Leu  
 115 120 125

Glu Arg Gln Tyr Gln Leu Arg Leu Phe Lys Asp Gly Lys Leu Lys Tyr  
 130 135 140

NLEE01001W00.ST25.txt

Gln Val Leu Asp Gly Glu Met Tyr Pro Pro Ser Val Glu Glu Ala Pro  
 145 150 155 160

Val Leu Met His Tyr Pro Arg Gly Ile Pro Pro Gln Ser Gln Met Ala  
 165 170 175

Val Gly Gln Glu Val Phe Gly Leu Leu Pro Gly Leu Met Leu Tyr Ala  
 180 185 190

Thr Leu Trp Leu Arg Glu His Asn Arg Val Cys Asp Leu Leu Lys Ala  
 195 200 205

Glu His Pro Thr Trp Gly Asp Glu Gln Leu Phe Gln Thr Thr Arg Leu  
 210 215 220

Ile Leu Ile Gly Glu Thr Ile Lys Ile Val Ile Glu Glu Tyr Val Gln  
 225 230 235 240

Gln Leu Ser Gly Tyr Phe Leu Gln Leu Lys Phe Asp Pro Glu Leu Leu  
 245 250 255

Phe Gly Val Gln Phe Gln Tyr Arg Asn Arg Ile Ala Met Glu Phe Asn  
 260 265 270

His Leu Tyr His Trp His Pro Leu Met Pro Asp Ser Phe Lys Val Gly  
 275 280 285

Ser Gln Glu Tyr Ser Tyr Glu Gln Phe Leu Phe Asn Thr Ser Met Leu  
 290 295 300

Val Asp Tyr Gly Val Glu Ala Leu Val Asp Ala Phe Ser Arg Gln Ile  
 305 310 315 320

Ala Gly Arg Ile Gly Gly Gly Arg Asn Met Asp His His Ile Leu His  
 325 330 335

Val Ala Val Asp Val Ile Arg Glu Ser Arg Glu Met Arg Leu Gln Pro  
 340 345 350

Phe Asn Glu Tyr Arg Lys Arg Phe Gly Met Lys Pro Tyr Thr Ser Phe  
 355 360 365

Gln Glu Leu Val Gly Glu Lys Glu Met Ala Ala Glu Leu Glu Glu Leu  
 370 375 380

Tyr Gly Asp Ile Asp Ala Leu Glu Phe Tyr Pro Gly Leu Leu Leu Glu  
 385 390 395 400

NLEE01001WO0.ST25.txt

Lys Cys His Pro Asn Ser Ile Phe Gly Glu Ser Met Ile Glu Ile Gly  
 405 410 415

Ala Pro Phe Ser Leu Lys Gly Leu Leu Gly Asn Pro Ile Cys Ser Pro  
 420 425 430

Glu Tyr Trp Lys Pro Ser Thr Phe Gly Gly Glu Val Gly Phe Asn Ile  
 435 440 445

Val Lys Thr Ala Thr Leu Lys Lys Leu Val Cys Leu Asn Thr Lys Thr  
 450 455 460

Cys Pro Tyr Val Ser Phe Arg Val Pro Asp Ala Ser Gln Asp Asp Gly  
 465 470 475 480

Pro Ala Val Glu Arg Pro Ser Thr Glu  
 485

<210> 32  
 <211> 122  
 <212> PRT  
 <213> HUMAN

<400> 32

Met Lys Leu Leu Thr Gly Leu Val Phe Cys Ser Leu Val Leu Gly Val  
 1 5 10 15

Ser Ser Arg Ser Phe Phe Ser Phe Leu Gly Glu Ala Phe Asp Gly Ala  
 20 25 30

Arg Asp Met Trp Arg Ala Tyr Ser Asp Met Arg Glu Ala Asn Tyr Ile  
 35 40 45

Gly Ser Asp Lys Tyr Phe His Ala Arg Gly Asn Tyr Asp Ala Ala Lys  
 50 55 60

Arg Gly Pro Gly Gly Val Trp Ala Ala Glu Ala Ile Ser Asp Ala Arg  
 65 70 75 80

Glu Asn Ile Gln Arg Phe Phe Gly His Gly Ala Glu Asp Ser Leu Ala  
 85 90 95

Asp Gln Ala Ala Asn Glu Trp Gly Arg Ser Gly Lys Asp Pro Asn His  
 100 105 110

Phe Arg Pro Ala Gly Leu Pro Glu Lys Tyr  
 115 120

## NLEE01001W00.ST25.txt

<210> 33  
<211> 26  
<212> DNA  
<213> HUMAN  
  
<400> 33  
agatattgca cgggagaata tacaaa 26  
  
<210> 34  
<211> 27  
<212> DNA  
<213> HUMAN  
  
<400> 34  
tcaattcctg aaattaaagt tcggata 27  
  
<210> 35  
<211> 23  
<212> DNA  
<213> HUMAN  
  
<400> 35  
tctgcagagt tggaagcact cta 23  
  
<210> 36  
<211> 21  
<212> DNA  
<213> HUMAN  
  
<400> 36  
gccgaggctt ttctaccaga a 21  
  
<210> 37  
<211> 20  
<212> DNA  
<213> HUMAN  
  
<400> 37  
catggcttga tcagcaagga 20  
  
<210> 38  
<211> 21  
<212> DNA  
<213> HUMAN  
  
<400> 38  
tggaagtgtg ccctgaagaa g 21  
  
<210> 39  
<211> 21  
<212> DNA  
<213> HUMAN  
  
<400> 39  
aagcagcacc agcaagtga g 21

NLEE01001WO0.ST25.txt

<210> 40  
<211> 21  
<212> DNA  
<213> HUMAN

<400> 40  
tcatggcctg tgtcagtcaa a 21

<210> 41  
<211> 22  
<212> DNA  
<213> HUMAN

<400> 41  
acatgccagc cactgtgata ga 22

<210> 42  
<211> 21  
<212> DNA  
<213> HUMAN

<400> 42  
ccctgccttc acaatgatct c 21

<210> 43  
<211> 23  
<212> DNA  
<213> HUMAN

<400> 43  
ggaattcacc tcaagaacat cca 23

<210> 44  
<211> 23  
<212> DNA  
<213> HUMAN

<400> 44  
agtgtggcta tgacttcggt ttg 23

<210> 45  
<211> 22  
<212> DNA  
<213> HUMAN

<400> 45  
cagccacaag cagtccagat ta 22

<210> 46  
<211> 24  
<212> DNA  
<213> HUMAN

<400> 46  
cctgactatc aatcacatcg gaat 24

NLEE01001WO0.ST25.txt

<210> 47  
<211> 21  
<212> DNA  
<213> HUMAN  
  
<400> 47  
ccaggtgctc cacatgacag t 21  
  
<210> 48  
<211> 24  
<212> DNA  
<213> HUMAN  
  
<400> 48  
aaacaaccaa caacaaggag aatg 24  
  
<210> 49  
<211> 21  
<212> DNA  
<213> HUMAN  
  
<400> 49  
cgtctccaca catcagcaca a 21  
  
<210> 50  
<211> 22  
<212> DNA  
<213> HUMAN  
  
<400> 50  
tcttggcagc aggatagtcc tt 22  
  
<210> 51  
<211> 22  
<212> DNA  
<213> HUMAN  
  
<400> 51  
gcagaccagc atgacagatt tc 22  
  
<210> 52  
<211> 20  
<212> DNA  
<213> HUMAN  
  
<400> 52  
gcggattagg gcttcctctt 20  
  
<210> 53  
<211> 23  
<212> DNA  
<213> HUMAN  
  
<400> 53  
tgaagttcaa tgcactggaa ctg 23

NLEE01001wo0.ST25.txt

<210> 54  
<211> 20  
<212> DNA  
<213> HUMAN  
  
<400> 54  
caggacgatc tccacagcaa 20  
  
<210> 55  
<211> 23  
<212> DNA  
<213> HUMAN  
  
<400> 55  
tggagtccac gagatcattt aca 23  
  
<210> 56  
<211> 19  
<212> DNA  
<213> HUMAN  
  
<400> 56  
agccttg gcc ctcggatat 19  
  
<210> 57  
<211> 21  
<212> DNA  
<213> HUMAN  
  
<400> 57  
cactgagttc gccaagagca t 21  
  
<210> 58  
<211> 23  
<212> DNA  
<213> HUMAN  
  
<400> 58  
cacgccatac ttgagaaggg taa 23  
  
<210> 59  
<211> 23  
<212> DNA  
<213> HUMAN  
  
<400> 59  
gctagtgatc aacagtggca atg 23  
  
<210> 60  
<211> 18  
<212> DNA  
<213> HUMAN  
  
<400> 60  
gctggcctct ccgttgag 18

NLEE01001wo0.ST25.txt

<210> 61  
<211> 22  
<212> DNA  
<213> HUMAN

<400> 61  
tgttcgggtg ccagttccaa ta

22

<210> 62  
<211> 22  
<212> DNA  
<213> HUMAN

<400> 62  
tgccagtggg agagatgggt ga

22

<210> 63  
<211> 22  
<212> DNA  
<213> HUMAN

<400> 63  
gggacatgtg gagagcctac tc

22

<210> 64  
<211> 21  
<212> DNA  
<213> HUMAN

<400> 64  
catcatagtt ccccccagca t

21